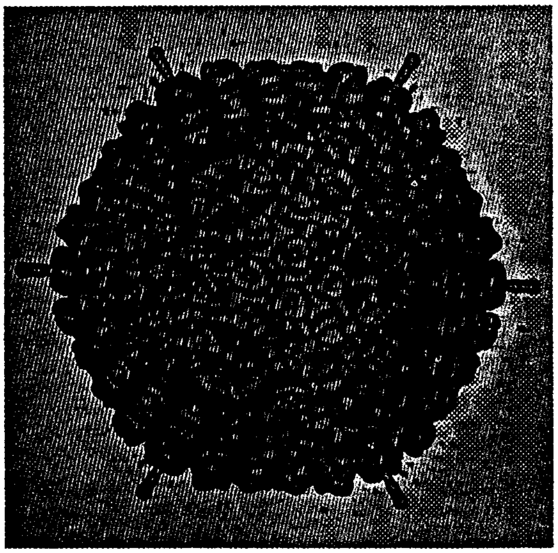
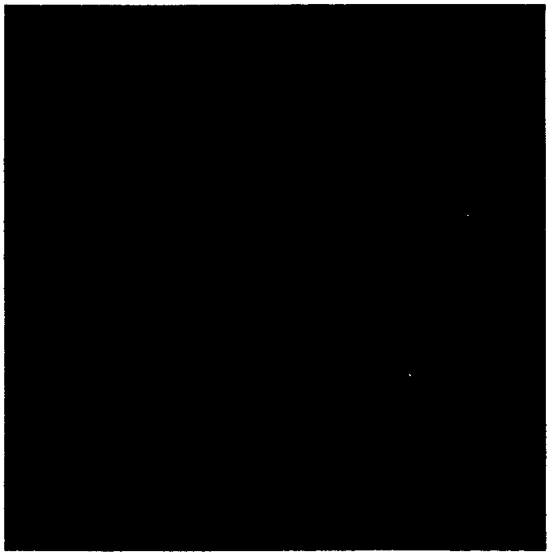
PROCEEDINGS

VISUALIZATION '90

OCTOBER 23-26, 1990
SAN FRANCISCO, CALIFORNIA
EDITED BY ARIE KAUFMAN





SPONSORED BY THE IEEE COMPUTER SOCIETY TECHNICAL COMMITTEE ON COMPUTER GRAPHICS, ACM and SIGGRAPH

IEEE Computer Society Press

The Institute of Electrical and Electronics Engineers, Inc.

A System for Three-Dimensional Acoustic "Visualization" in a Virtual Environment Workstation

Elizabeth M. Wenzel
NASA-Ames Research Center, MS 262-2
Moffett Field, CA 94035

Philip K. Stone
Sterling Software
NASA-Ames Research Center, MS 262-6
Moffett Field, CA 94035

Scott S. Fisher
11571 Buena Vista Drive
Los Altos Hills, CA 94022

Scott H. Foster
Crystal River Engineering
12350 Wards Ferry Road
Groveland, CA 95321

Abstract

This paper describes the real time acoustic display capabilities developed for the VIrtual Environment Workstation (VIEW) Project at NASA-Ames Research Center. The acoustic display is capable of generating localized acoustic cues in real time over headphones. An auditory symbology, a related collection of representational auditory "objects" or "icons," can be designed using ACE, the Auditory Cue Editor, which links both discrete and continuously-varying acoustic parameters with information or events in the display. During a given display scenario, the symbology can be dynamically co-ordinated in real time with three-dimensional visual objects, speech, and gestural displays. The types of displays feasible with the system range from simple warnings and alarms to the acoustic representation of multidimensional data or events.

Introduction

Recent years have seen many advances in computing technology with the associated requirement that users manage and interpret increasingly complex systems of information. As a result, an increasing amount of applied research has been devoted to a type of reconfigurable interface called the virtual display. Some of the earliest work in this area was done by Sutherland [30] at the University of Utah using binocular head-mounted displays. Sutherland characterized the goal of virtual interface research, stating that, "The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real." As the technology has advanced, virtual displays have gone beyond the flat CRT screen, assuming a three-dimensional spatial organization which, it is hoped, provides a richer and more natural means of accessing and manipulating information. A few projects have taken the spatial metaphor to its limit by directly involving the operator in the data environment [5], [19], [21]. Thus, the kind of "artificial reality" once relegated solely to the specialized world of the cockpit simulator is now being seen as a next step in interface development for all types of advanced computing applications [20].

Auditory Icons & Symbologies

As with most research in information displays, virtual displays have generally emphasized visual information. Many investigators, however, have pointed out the importance of the auditory system as an alternative or supplementary information channel, particularly when the visual channel is overloaded and visual cues are degraded or absent [12], [13], [27]. Most recently, attention has been devoted to the use of non-speech audio as an interface medium [1], [2], [8], [23]. Auditory signals are detected more quickly than visual signals and tend to produce an alerting or orienting response. Consequently, non-speech audio has been most frequently used in simple alarm or warning systems, as in aircraft cockpits or the siren of an ambulance. Another advantage of audition is that it is primarily a temporal sense and we are extremely sensitive to changes in an acoustic signal over time. This feature tends to bring any such acoustical event to our attention and conversely, allows us to relegate sustained or uninformative sounds to the background. Thus audio is particularly suited to monitoring changes over time, for example when your car engine suddenly begins to malfunction. Non-speech signals have the potential to provide an even richer display medium if they are carefully designed with human perceptual abilities in mind. Just as a movie with sound is much more compelling and informationally-rich than a silent film, so could a computer interface be enhanced by an appropriate "sound track" to the task at hand. If used properly, sound need not be distracting or cacophonous or merely uninformative. Principles of design for auditory icons and symbologies can be gleaned from the fields of music, psychoacoustics, and psychological studies of the acoustical determinants of perceptual organization. For example, one can think of the audible world as being composed of a collection of acoustic "objects." Various acoustic features, such as timbre, intensity, and temporal rhythm, specify the

CH2913-2/90/0000/0329/$01.00 – 1990 IEEE

identities of the objects and perhaps convey meaning about discrete events or ongoing actions in the world and their relationships to one another. One could systematically manipulate these features and create an auditory symbology which operates on a continuum from "literal" everyday sounds, such as the clunk of mail in your mailbox (e.g., Gaver's 'Sonic Finder' [23]), to a completely abstract mapping of statistical data into sound parameters [4], [28].

Such a display could be further enhanced by taking advantage of the auditory system's ability to segregate, monitor, and switch attention among simultaneous sources of sound. One of the most important determinants of acoustic segregation is an object's location in space.

A true three-dimensional auditory display could potentially improve information transfer by combining directional and iconic information in a quite naturalistic representation of dynamic objects in the interface. Borrowing a term from Gaver [23], an obvious aspect of "everyday listening" is the fact that we live and listen in a three-dimensional world. Indeed, a primary advantage of the auditory system is that it allows us to monitor and identify sources of information from all possible locations, not just the direction of gaze. This feature would be especially useful in an application that is inherently spatial, such as an air traffic control display for the tower or cockpit, or even in a two-dimensional interface which has adopted a spatial organization, such as the desktop metaphor. A further advantage of the binaural system, often referred to as the "cocktail party effect" [10], [16], is that it improves the intelligibility of sources in noise and enhances the segregation of multiple sound sources. This effect could be critical in applications involving encoded information as in scientific "visualization," using the acoustic representation of multi-dimensional data [4], [28], or the development of alternative interfaces for the visually impaired [15], [28]. Another aspect of auditory spatial cues is that, in conjunction with other modalities, they can act as a potentiator of information in the display. That is, visual and auditory cues together can reinforce the information content of the display and provide a greater sense of presence or realism in a manner not readily achievable by either modality alone [1], [7], [11], [26], [29], [31]. This phenomenon will be particularly useful in telepresence applications, such as advanced teleconferencing environments, shared electronic workspaces, or monitoring telerobotic activities in remote or hazardous situations. Thus, the combination of direct spatial cues with good principles of iconic design could provide an extremely powerful and information-rich display which is also quite easy to use.

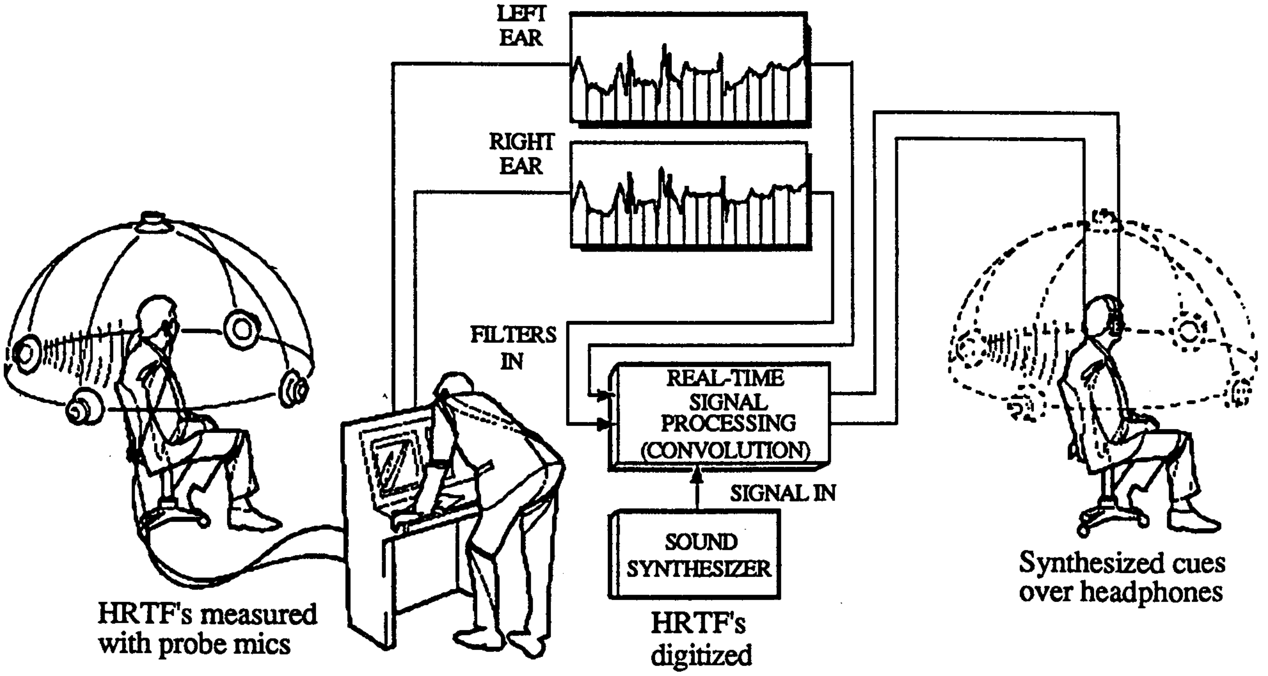
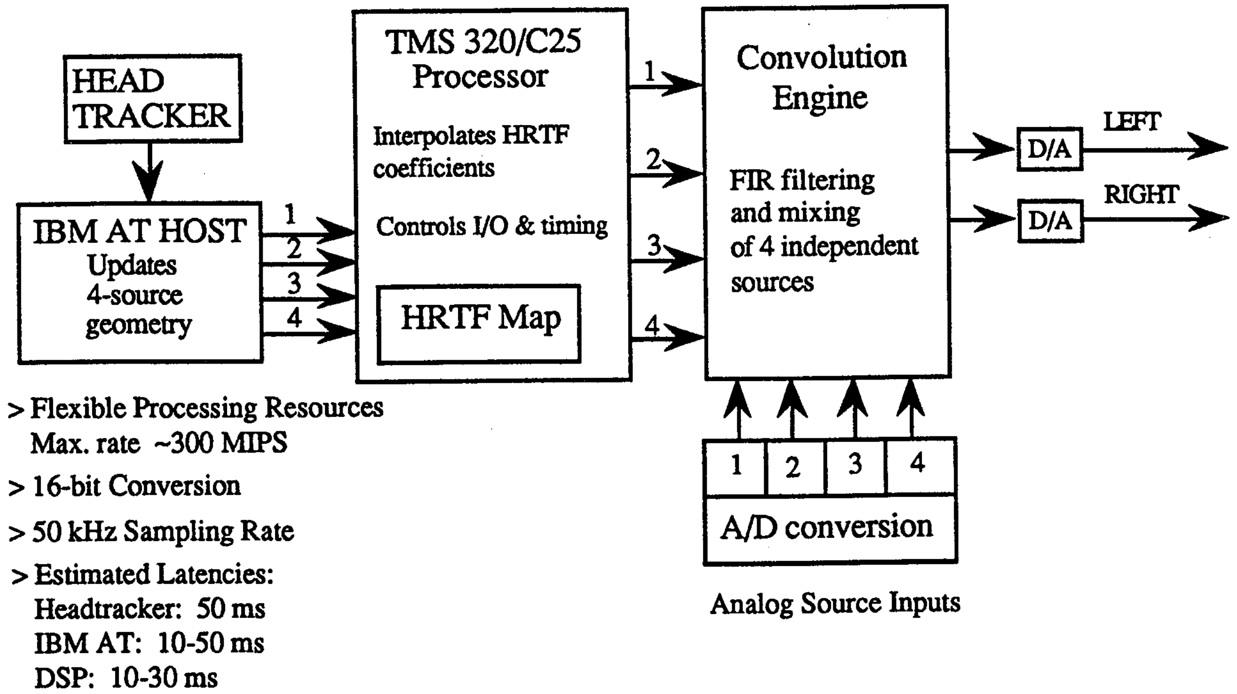

LEFT EAR
RIGHT EAR
FILTERS IN
REAL-TIME SIGNAL PROCESSING (CONVOLUTION)
SIGNAL IN
SOUND SYNTHESIZER
HRTF's measured with probe mics
HRTF's digitized
Synthesized cues over headphones

Figure 1: 3D Auditory Display; Synthesis Technique



> Flexible Processing Resources
Max. rate ~300 MIPS

> 16-bit Conversion

> 50 kHz Sampling Rate

> Estimated Latencies:
Headtracker: 50 ms
IBM AT: 10-50 ms
DSP: 10-30 ms

Fig. 2: The Convolvotron

Implementing Three-Dimensional Sound

Localized acoustic cues could be realized with an array of real sound sources or loudspeakers [9], [13]. An alternative approach, recently developed at NASA-Ames Research Center, generates externalized, three-dimensional sound cues over headphones in real time using digital means [32], [33]. This type of presentation system is desirable because it allows complete control over the acoustic waveforms delivered to the two ears and the ability to interact dynamically with the virtual display. The synthesis technique, illustrated in Figure 1, involves the digital generation of stimuli using Head-Related Transfer Functions (HRTFs) measured in the ear-canals of individual subjects (see [36], [3]). The advantage of this technique is that it preserves all of the interaural temporal and level differences over the entire spectrum of the stimulus, thus capturing the effects of filtering by the pinnae which are critical for the veridical simulation of externalized sound sources.

In the real time system, the Convolvotron, up to four moving or static sources can be simulated in a head-stable environment by digital filtering of arbitrary signals with the appropriate HRTFs. Motion trajectories and static locations at greater resolutions than the empirical data are simulated by linear interpolation of the four nearest measured transforms. Also, a simple distance cue is provided via real time scaling of amplitude. Figure 2 shows the functional components of the Convolvotron system designed by Scott Foster.

Such an interface not only requires the development of special-purpose display technology, it also necessitates the careful psychophysical evaluation of listeners' ability to accurately localize the virtual or synthetic sound sources. The working assumption of the synthesis technique is that if, using headphones, one can produce ear-canal waveforms identical to those produced by a free-field source, the free-field experience will be duplicated. A recent study [36] confirmed the perceptual adequacy of the basic technique for static sources for experienced subjects localizing stimuli in the free-field compared with stimuli synthesized from their own HRTFs. Source azimuth was synthesized nearly perfectly for all listeners while synthesis of source elevation was less well-defined, e.g., more variable with a compressed range of responses. Elevation was also the source of the most obvious individual differences in localization for both free-field and synthesized signals.

Unfortunately, measurement of each potential listener's HRTFs may not be possible in practice. It may also be the case that the user will not have the opportunity for extensive training. Thus, a critical research issue for virtual acoustic displays is the degree to which the general population of listeners can obtain adequate localization cues from stimuli based on non-individualized transforms. Preliminary data [34] from

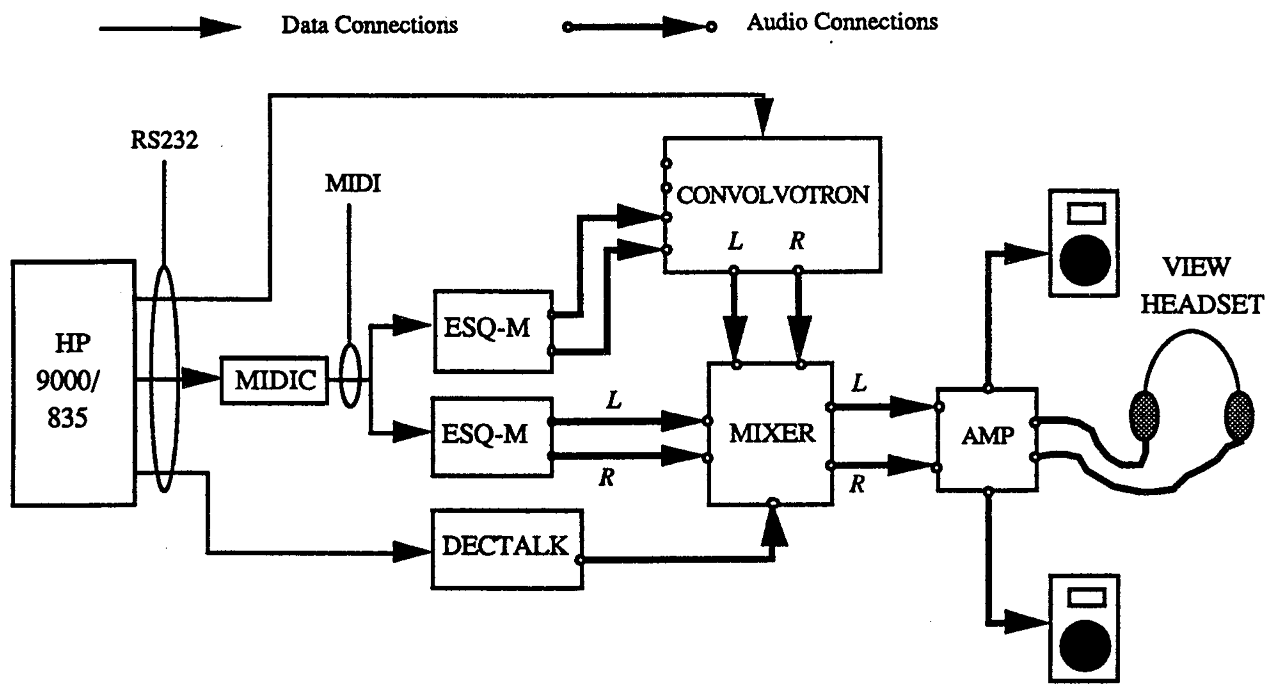


Figure 3: VIEW Auditory Display; System Overview

three experienced subjects suggest that using non-listener-specific transforms to achieve synthesis of localized cues is at least feasible. Localization performance was only somewhat degraded compared to a subject's inherent ability, even for the less robust elevation cues, as long as the transforms are derived from what one might call a "good" localizer. Further, the fact that individual differences in performance, particularly for elevation, can be traced to acoustical idiosyncrasies in the HRTF spectra, suggests that it may eventually be possible to create a set of "universal transforms" by parametric modeling techniques (e.g., [24]), principal components analysis, or perhaps even enhancing the spectra of empirically-derived transfer functions (e.g., [14]).

VIEW Sound System Architecture

While perceptual studies of individual sensory modalities are clearly needed, it is also important to examine the role of sensory interaction. NASA-Ames' VIEW system provides the opportunity to implement localized auditory icons and assess their contribution to an integrated spatial display. Briefly, VIEW is a multisensory display environment which allows the user to explore and interact with a 360-degree synthesized, or remotely-sensed, world using a head-mounted, wide-angle stereoscopic display controlled by operator position, voice, and gesture. More detailed descriptions of the VIEW visual and gestural displays can be found in [17] - [19].

The VIEW auditory display subsystem allows audio cues responsive to both discrete events and continuous data changes to be designed and linked to arbitrary events and data flows in VIEW scenarios. Refer to the system overview diagram, Figure 3, for the following discussion.

Development of the initial binaural display capability based on MIDI (Musical Instrument Digital Interface) sound-synthesis technology began in 1987. More recently, true spatial cueing was added to the system with the integration of the Convolvotron. The auditory display subsystem, like most of the VIEW system, is currently implemented with a Hewlett Packard HP9000/835 computer. Two Ensoniq ESQ-M synthesizer modules handle the actual production of audio cues, supplemented with a Digital Equipment Corporation DECTalk speech synthesizer. MIDI protocol is used to communicate between the HP host and the ESQ synthesizers. A Hinton Instruments MIDIC interface converts 19.2 Kbps RS232 signals from the HP into 31.25 Kbps 5 mA. current-loop signals that are required by the MIDI standard.

Each ESQ synthesizer has two outputs. One ESQ's output pair is mapped directly into the VIEW system's left and right audio channels. Up to eight independent (polytimbral) voices, mixed to the stereo output, may be played through this synthesizer. The second ESQ's output pair is patched into the Convolvotron. As described above, this device is capable of synthesizing, in real time, an apparent three-dimensional location for up to four independent audio inputs. In this case, since only two channels of sound are available from the ESQ, only two of the Convolvotron's inputs are used. The simple stereo pair from the first ESQ, and the 3D-imaged output from the Convolvotron, are mixed, amplified, and sent to headphones integrated into the VIEW headset, or optionally, to room speakers.

The central software component of the auditory display is the cue driver. As in all VIEW applications, the software is written in C in a Unix environment. Without delving too deeply into its implementation details, it can be described as consisting of an event scheduler, a MIDI, speech, and Convolvotron event generator, and a VIEW/Auditory Display rendezvous mechanism. It also handles several housekeeping chores, such as loading cue files, initializing the MIDI interface, the synthesizers and the Convolvotron, downloading patch files to the synthesizers, and making sure all is quiet before a VIEW scenario exits.

Up to ten (monophonic) auditory "objects" may be displayed simultaneously. If a more complex, or polyphonic, sound is desired, several voices can be assigned to the same icon and the number of possible simultaneous objects is reduced accordingly. In general, the basic sound signature or identity of an individual object or "icon" derives from the particular ESQ patch assigned to it. Since this technology was developed for music synthesis, one can often think of a patch as having the attributes of a particular musical instrument. However, some "environmental" sounds analogous to sound effects, e.g., footsteps or explosions, are also possible.

Custom ESQ patches may be designed off-line using the front-panel capabilities of the synthesizer, a patch editor/librarian software package, or by selecting from collections of commercially-available pre-programmed patches. Icons may take advantage of one or more of the controllable parameters made available by the ESQ. These include oscillator frequency, filter cutoff frequency, amplitude level, and stereo pan position. Any of these can be modulated in real time and associated with events or information flow in a VIEW display scenario. One of the advantages of the ESQ, and the main reason for its choice during the specification of this system, is that it allows access to these parameters through standard MIDI controllers; many synthesizers require the use of "system-exclusive" messages to achieve this level of control. Designing the cue driver around standard MIDI controllers makes it less system-dependent; as more synthesizers adopt this level of control (and this seems to be the trend), the auditory display may be readily adapted to them.

Because of the limited outputs of the ESQ (two per synthesizer), only two of the ten icons may be assigned specific locations via the Convolvotron at any one time. Alternatively, using all four ESQ outputs, up to four simultaneous icons could be independently localized. The current configuration was a compromise solution which traded localized cues for an increase in the number of possible icons. A future solution would be to adapt the system to a synthesizer which has independent outputs for each voice. Also, integrating a digital-sampling device would be useful for presenting the kinds of sounds that Gaver [23] advocates in his notion of "everyday listening." At the time we began developing the system, digital samplers tended to be expensive and allowed very little real time control over the acoustic parameters of the sounds. Since a major goal of the display was to allow continuous control over the icons' acoustic structure, we opted for a more standard and inexpensive synthesizer with a relatively well-developed MIDI implementation. [See [8] for a useful discussion of the pros and cons of various MIDI devices.]

Editing & Display Capabilities

Cues or icons are designed and refined with ACE, the Auditory Cue Editor. ACE is a stand-alone program, which makes it unnecessary to activate the entire VIEW system merely to work on auditory cues. Multi-level menus, interactive prompting, and extensive syntax checking aid the user in designing complex auditory cues with relative ease.

ACE is composed of four basic sections, organized as independent screens, each with its own menu of commands.

On the Main Screen, cues may be created, deleted, loaded, saved, and named (an important function; all rendezvous between VIEW events and data are made through the names of cues as specified on the Main Screen). The Main Screen may also be used to specify certain basic parameters, e.g., synthesizer patch number and localization method (convolved or simple stereo). In addition, a "play" command allows quick, interactive audition of cues during their design.

The Sound Event Editing Screen allows the construction of the main body of a cue, which is in the form of a list of time-ordered MIDI, speech, or Convolvotron localization commands. While entire pieces of music could theoretically be entered here, note by painstaking note, this is usually not the case. Single notes, chords, or short sequences of notes and chords are the most common items entered on this screen. This is because very simple events (with carefully chosen and distinct timbres) are often all that is needed for basic cueing functions. Even the more complex auditory icons generally have fairly simple event lists, since most higher-level display capabilities result from linkage to continuous data streams and response to real-time changes of those data streams.

The Modulation Editing Screen enables the connection and scaling of incoming data streams to several different MIDI modulators as well as Convolvotron sound-source position coordinates. This makes it possible to have an arbitrary data value produced by the VIEW system displayed as a proportional deflection in a variety of auditory parameters, such as pitch, timbre, or apparent three-dimensional location. The data structure has provisions for incorporating nonlinear mappings between incoming data and an auditory parameter, a feature which can be very important in evaluating the perceptual consequences of a particular icon. Specifications of modulation behavior composed on this screen are given textual names, which enables rendezvous with the appropriate VIEW-generated data streams.

Patches, as mentioned above, are a very important part of an auditory icon; they define the basic sound that the synthesizer produces, and the manner in which it will react to incoming controls sent by the cue driver. The Patch Maintenance Screen allows the uploading and downloading of individual patches and complete patch banks between the host system and the synthesizers. In this way, a particular set of sound programs can be directly associated with a set of auditory cue definitions.

Application in the Virtual Environment: Telerobotic Control

The VIEW telerobotic scenario was designed to illustrate the capability of telepresence, i.e., the manipulation of objects or interaction with persons or objects remote from one's location, that a virtual environment makes possible. In this scenario, the visual and kinematic characteristics of a Puma robotic arm are modeled with a high degree of accuracy. The scenario participant may, upon donning the VIEW stereoscopic head-mounted display, align his or her arm with that of the lifesize model. The participant's arm and the modeled robotic arm may then be "coupled," which simply means that the robotic arm will move, to the extent of its kinematic capabilities, in correspondence to the movement of the participant's arm. During the coupled mode, an end-effector with a vise-like gripping apparatus may be opened and closed merely by opening and closing the hand.

This graphic computational model of a robotic arm is meant to test the efficacy of the intuitive mapping of control between machine and human counterpart. The success of this mapping is tested by assigning a simple task to be completed by the telerobotic participant. In the foreground of the scenario, a "circuit card" is plugged into a slot on a "task board." The participant is instructed to remove the circuit card and replace it with another one which is just off to the side of the task board. This entails coupling with the robotic arm, maneuvering the end-effector into position so that the two jaws surround the edge of the circuit card, closing the end-effector jaws around the card to grasp it, and pulling it out and away from the task board. Once this is completed, the replacement circuit card must be grasped in a similar manner, lined up exactly, and inserted into the slot.

With perfect telepresence, this task could be accomplished with little more difficulty than if one were using one's own hand and a real task board and circuit card. However, factors such as slower-than-ideal graphic refresh rates, lower-than-ideal contrast and focus in the VIEW display, etc., conspire to make the precision manipulation required somewhat difficult. In a situation such as this, auditory feedback can make an important difference, particularly with the current paucity of good haptic or force feedback display systems.

At the simplest level, auditory feedback is used to indicate the occurrence of discrete events in the scenario. For example, many commands and actions in VIEW are initiated by hand gesture. A VPL "Data Glove" reports finger positions to the host computer, which examines those positions for correspondence to any of several pre-defined gestures, such as "single-finger point," or "fist." When one of these gestures is detected by the host, a sound is made by the auditory display to indicate gesture recognition.

Other simple auditory cues fall into the category of "reality-mimicry," or sound effects. In the telerobotic scenario, bumping an end-effector into a "solid" object in the virtual (or teleoperated) world causes a "bump" sound to be produced. Since direct force feedback is not yet available in the VIEW system, this form of audio display is particularly critical, as it warns of a situation which could cause damage to a real-world robotic arm or to objects with which it is colliding. At a more mundane level, this sort of sound effect enhances the sense of presence; objects tend to make a sound when they collide in the real world, so it is reasonable to expect them to do so in a virtual environment.

Audio feedback that supplements or replaces force feedback is not limited to mimicry of collision sounds or similar sound effects. Force can be represented as a continuum by changing one or more sound parameters in correspondence to the force's intensity. This type of display is utilized for a special circumstance in the telerobotic scenario. If the scenario participant attempts to force the replacement circuit card into the task board without orienting it correctly, a force-reflection display called "push-through" is initiated. It starts out as a soft, steady tone that gets louder, brighter (higher harmonics are let through the filter), and more frequency modulated the harder the participant pushes on the misaligned card. In this way, a potentially damaging increase in user input is signalled by an increasingly harsh and strident auditory warning.

Taking this idea one step further, not only force, but any arbitrary continua of data may be displayed. Perhaps the most successful use of auditory feedback in the telerobotic scenario comes into play while the participant attempts to guide the replacement circuit board into the target socket. As the board reaches a certain proximity to the socket, a cue initiates consisting of two sustained tones; the pitch of one of the tones is deflected with respect to the other by an amount proportional to the distance between the circuit card and its slot. As the card nears the slot, the two pitches come closer together (which is readily perceived due to the obvious decrease in the beat frequency produced by the increasingly adjacent pitches); at distance zero, the tones are in unison. The cue functions as an auditory "rangefinder," and greatly facilitates the proper positioning of the card in its slot.

Card orientation, which is also crucial to the completion of the replacement task, could be represented by some other continuous audio parameter, such as depth of frequency modulation. With careful selection and scaling of the modulated sound parameters, two continua (e.g., proximity and orientation) could be monitored simultaneously at a very intuitive level. Our work to date has not explored multiple-simultaneous displays of continuous data, but it is an intriguing area for further research.

Auditory Design Principles

The telerobotic scenario has served as an excellent test of the capabilities of the auditory display. While formal experimentation has yet to be done, it has also provided a rich environment for the discovery of certain basic guidelines for the design of auditory icons and the development of an auditory symbology.

Practical experience has shown that the most effective cues are simple cues. Long sequences or elaborate clusters of tones not only tend to clutter the auditory display, they can increase the load on cognitive processing and memory required to interpret the information, and in the long run, become downright annoying. Imagine a telephone that played "Three Blind Mice" every time a call came in. It would be only a short time before the exactly-repeating melody became maddening. The simple bell or digital chirp of a telephone manages to get attention without engaging the "music critic" part of one's cognition. Similarly, a "thud" sound suffices to signal a bump in a virtual world; it is not necessary to have a speech synthesizer say "You have bumped into something" at each and every collision. While these may be extreme examples, the basic principle holds; an auditory icon should be as simple as possible.

The need for simplicity is even more critical when several cues occur in close proximity to each other. For instance, in the telerobotic scenario, a gesture-recognition cue might be followed immediately by a sound that indicates movement of the jaws of the end-effector. If the jaws were then to close over the circuit board, a "board-grasped" cue would result. These three cues can occur in rapid succession, so they must be of short duration for the correct sequence of events to be properly represented.

This situation also points out the need for carefully choosing the sound signatures or patches which form the fundamental units of an auditory symbology. Patch design, including spectral content, amplitude and filter envelopes and various special effects, is the chief distinguishing feature of a simple icon. The best way to make an icon recognizable is to give it a distinctive sound. Much effort in the design of auditory icons is therefore concentrated in selecting or building an appropriate synthesizer patch.

As noted before, guidelines can be derived from the fields of psychoacoustics, music, and perceptual psychology. As illustrated in the proximity cue in the telerobot scenario, the close tuning of two pitches is a continuous parameter to which the human ear is very sensitive. However, the amplitude modulation or beat frequency which signals the change in proximity will only occur for a limited range of frequencies which must be considered when mapping the distance data to the difference in pitch. In developing a symbology, one can also take advantage of what one might think of as "natural" or metaphorical mappings even if a literal sound, such as a "bump," is not possible. For example, the "push-through" cue described above clearly signals an increasing violation of the allowable forward movement when inserting the task board at an incorrect orientation by a harsh sound which increasingly "violates" the ears. To minimize cognitive effort, it is important to build meaning by the relationships between icons as well. In the telerobot scenario, icons which provide feedback for related gestures have similar timbres that are distinguished by their temporal structure. For example, larger changes in pitch, at the level of short sequences, are used (much like the familiar two-chime doorbell). This particular type of icon has the virtue of being reversible like a short musical motive; a "grasp" gesture is represented by a high note followed by a slightly lower note. The complementary "release" gesture is the same two notes, only in reverse order. As much as possible, this relationship between sound and meaning should remain consistent throughout a display system. Thus, in VIEW, the cues which provide feedback for the various gestures remain the same across the different types of display scenarios that have been developed.

Careful consideration of the possible interactions between icons will be particularly important when auditory cues must be presented simultaneously as in the combination of orientation and proximity cueing described above. Principles of acoustic perceptual organization, such as the Gestalt principles elaborated by Bregman [6], will provide important guidelines. For example, different acoustic objects may be defined by different auditory streams. Streaming is determined by such features as frequency separation, timbre, rate or tempo, spatial location, and "common fate" or the tendency of spectral components to be grouped according to similar frequency or temporal patterns.

Other VIEW Scenarios

Other examples of display scenarios which have been implemented in the VIEW system include the Extra-Vehicular Activity (EVA) Visor and a Computational Fluid Dynamics (CFD) data visualization. The EVA Visor is a concept for a helmet-mounted, three-dimensional dataspace which can be accessed by an astronaut during repair or inspection activities while outside the space station. In the scenario, several types of display windows can be used, including life-support system status, a "cuff checklist" of tasks to be completed, repair schematics, and a three-dimensional "map cube," which represents the entire EVA scenario as a miniature, manipulable cube, with which the astronaut can establish his or her position and orientation with respect to the other vehicles or objects in the scenario. Currently, auditory cues for the EVA Visor include a set of gesture recognition cues identical to those in the telerobot scenario (in keeping with the principle of transference of a learned auditory vocabulary between virtual worlds, when possible). Warning cues signal situations which might endanger the astronaut's safety, such as impending depletion of life support resources. A special sound effect cue indicates when MMU (the rocket-backpack vehicle which enables the astronaut to maneuver during EVA) thrusters are firing. Finally, activation of the various windows listed above is heralded by corresponding audio signals. Another cue, not yet implemented, which would be useful in the EVA display, is an orientation beacon which allows the astronaut to continuously monitor the location of the space station by means of a localized auditory icon to minimize disorientation in the absence of visual and gravitational referents.

The CFD data display visualizes the fuel-flow around the LOX (liquid oxygen) post of the main shuttle engine. Features include the ability to "fly through" the data, viewing it from different viewpoints including inside the fluid flow, and "grabbing" and scaling the data up or down to examine its finer or coarser features. Although not yet implemented, a potentially useful auditory visualization cue might be to "attach" auditory icons to one or more particles in the flow, and thus follow their progress as they interact with the structures of the shuttle engine.

In developing the VIEW auditory display, we have attempted to provide a flexible and general-purpose system which takes advantage of our knowledge of perceptual abilities as much as possible. The hardware architecture and software is designed to be applicable to a wide variety of display configurations and to allow a consistent approach to the design of auditory symbologies based on knowledge gleaned from music, psychoacoustics, and perceptual psychology.

References

[1] Begault, D.R. & Wenzel, E.M. (1990) Techniques and applications for binaural sound manipulation in man-machine interfaces. NASA Technical Memorandum No. TM102279, In Press.

[2] Blattner, M.M., Sumikawa, D.A., & Greenberg, R.M. (1989) Earcons and icons: Their structure and common design principles. Hum.-Comp. Interact., 4, 11-44.

[3] Blauert, J. (1983) Spatial Hearing. The MIT Press: Cambridge, MA.

[4] Bly, S. (1982) Sound and computer information presentation. Unpublished doctoral thesis (UCRL-53282) Lawrence Livermore National Laboratory and University of California, Davis, CA.

[5] Brooks, F.P. (1988) Grasping reality through illusion -- Interactive graphics serving science. Proc. CHI'88, ACM Conf. Hum. Fac. Comp. Sys., Washington, D.C., 1-11.

[6] Bregman, A. (1981) Asking the "what for" question in auditory perception. In Kubovy & Pomerantz (Eds.), Perceptual Organization,Lawrence Erlbaum Associates: Hillsdale, NJ.

[7] Brown, M., Newsome, S., & Glinert, E. (1989) An experiment into the use of auditory cues to reduce visual workload. Proceedings of CHI'89, ACM Conference on Human Factors in Computing Systems, 339-346.

[8] Buxton, W., Gaver, W., & Bly, S. (1989) The use of non-speech audio at the interface. Tutorial #10, CHI'89, ACM Press: New York.

[9] Calhoun, G.L., Valencia, G., & Furness, T.A. III (1987) Three-dimensional auditory cue simulation for crew station design/evaluation. Proc. Hum. Fac. Soc., 31, 1398-1402.

[10] Cherry, E.C. (1953) Some experiments on the recognition of speech with one and two ears. J. Acoust. Soc. Am., 22, 61-62.

[11] Colquhoun, W.P. (1975) Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. Hum. Fac., 17, 425-437.

[12] Deatherage, B.H. (1972) Auditory and other sensory forms of information presentation. In H.P. Van Cott & R.G. Kincade (Eds.), Human Engineering Guide to Equipment Design, (rev. ed.), Washington, DC: U.S. Government Printing Office, 123-160.

[13] Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.

[14] Durlach, N.I. & Pang, X.D. (1986) Interaural magnification. J. Acoust. Soc. Am., 80, 1849-1850.

[15] Edwards, A.D.N. (1989) Soundtrack: An auditory interface for blind users. Hum. Comp. Interact.,
4, 45-66.

[16] Egan, J.P., Carterette, E.C., & Thwing, E.J. (1954) Some factors affecting multichannel listening. J. Acoust. Soc. Am., 26, 774-782.

[17] Fisher, S.S. (1986) Telepresence master glove controller for dexterous robotic end-effectors. Advances in Intelligent Robotics Systems, D.P. Casasent (Ed.), Proc. SPIE, 726.

[18] Fisher, S.S., McGreevy, M.W., Humphries, J., & Robinett, W. (1986) Virtual environment display system. ACM Workshop on Interactive 3D Graphics, Chapel Hill, NC.

[19] Fisher, S.S., Wenzel, E.M., Coler, C. & McGreevy, M.W. (1988) Virtual interface environment workstations. Proc. Hum. Fac. Soc., 32, 91-95.

[20] Foley, J.D. (1987) Interfaces for advanced computing. Sci. Amer., 257, 126-135.

[21] Furness, T.A. (1986) The super cockpit and its human factors challenges. Proc. Hum. Fac. Soc., 1986 (1), 48-52.

[22] Garner, W.R. (1949) Auditory signals. In A Survay Report on Human Factors in Undersea Warfare, Washington, D.C.: National Research Council, 201-217.

[23] Gaver, W. (1986) Auditory icons: Using sound in computer interfaces. Hum.-Comp. Interact., 2, 167-177.

[24] Genuit, K. (1986) A description of the human outer ear transfer function by elements of communication theory. Proc. 12th ICA (Toronto),Paper B6-8.

[25] Loomis, J.M., Hebert, C., & Cicinelli, J.G. (1990) Active localization of virtual sound sources. Submitted to J. Acoust. Soc. Am.

[26] O'Leary, A. & Rhodes, G. (1984) Cross-modal effects on visual and auditory object perception. Perc. & Psychophys., 35, 565-569.

[27] Patterson, R.R. (1982) Guidelines for Auditory Warning Systems on Civil Aircraft. Civil Aviation Authority Paper No. 82017, London.

[28] Smith, S., Bergeron, R.D., & Grinstein, G.G. (1990) Stereophonic and surface sound generation for exploratory data analysis. Proceedings of CHI'90, ACM Conference on Human Factors in Computing Systems, 125-132.

[29] Sorkin, R.D., Wightman, F.L., Kistler, D.J., & Elvers, G.C. (1989) An exploratory study of the use of movement-correlated cues in an auditory heads-up display. Hum. Fact., 31, 161-166.

[30] Sutherland, I.E. (1968) Head-mounted three-dimensional display. Proc. Fall Joint Comp. Conf., 33, 757-764.

[31] Warren, D.H., Welch, R., & McCarthy, T.J. (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. Perc. & Psychophys., 30, 557-564.

[32] Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988a) Development of a three-dimensional auditory display system. SIGCHI Bulletin,20, 52-57.

[33] Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988b) A virtual display system for conveying three-dimensional acoustic information. Proc. Hum. Fac. Soc., 32, 86-90.

[34] Wenzel, E.M., Wightman, F.L., Kistler, D.J., & Foster, S.H. (1988c) Acoustic origins of individual differences in sound localization behavior. J. Acoust. Soc. Amer., 84, S79.

[35] Wightman, F.L. & Kistler, D.J. (1989a) Headphone simulation of free-field listening I: stimulus synthesis. J. Acoust. Soc. Amer., 85, 858-867.

[36] Wightman, F.L. & Kistler, D.J. (1989b) Headphone simulation of free-field listening II: psychophysical validation. J. Acoust. Soc. Amer., 85, 868-878

COMPUTER GRAPHICS

Volume 24 • Number 2 • March 1990
A publication of ACM SIGGRAPH
Special issue on 1990 Symposium on Interactive 3D Graphics

acm PRESS

Realtime Digital Synthesis of Virtual Acoustic Environments

Elizabeth M. Wenzel
Mail Stop 239-3
NASA-Ames Research Center
Moffett Field, CA 94035

Scott H. Foster
Crystal River Engineering
12350 Wards Ferry Road
Groveland, CA 95321

As with most research in information displays, virtual displays have generally emphasized visual information. Many investigators, however, have pointed out the importance of the auditory system as an information channel. We believe that a three-dimensional auditory display can substantially enhance situational awareness by combining spatial and semantic information to form dynamic, multidimensional patterns of acoustic events which convey meaning about objects in the spatial world of the user. Such a display can be realized with an array of real sound sources or loudspeakers (Doll et. al., 1986). The signal-processing device being developed at NASA-Ames maximizes flexibility and portability by synthetically generating three-dimensional sound in realtime for delivery through headphones. Unlike conventional stereo, sources can be perceived outside the head at discrete distances and directions from the listener. The 3-D auditory display is currently being integrated with Ames' Virtual Interactive Environment Workstation (VIEW) which allows the user to explore and interact with a 360-degree synthesized or remotely-sensed world using a head-mounted, wide-angle, stereoscopic display controlled by operator position, voice, and gesture.

Applications of a three-dimensional auditory display involve any context in which the user's spatial awareness is important, particularly when visual cues are limited or absent. Examples include advanced teleconferencing environments, monitoring telerobotic activities in hazardous situations, and scientific "visualization" of multi-dimensional data. (e.g., Doll, et. al., 1986; Foley, 1987; Fisher, et. al., 1988; Brooks, 1988).

Synthesizing Out-of-Head Localization.

The synthesis technique is based on the Head-Related Transfer Function (HRTF); the listener-specific, direction-dependent acoustic effects imposed on an incoming signal by the outer ears. HRTFs in the form of Finite Impulse Responses (FIRs) are measured with small probe microphones placed near the two eardrums of a listener seated in an anechoic chamber for 144 different speaker locations at intervals of 15 degrees azimuth and 18 degrees elevation (range: -36 to +54) [see Wightman & Kistler, 1989a]. In order to synthesize localized sounds, a map of listener-specific "location filters" is constructed from the 144 pairs of FIR filters represented in the time domain. The map of FIR filters is downloaded from an IBM AT to the dual-port memory of a realtime digital signal-processor, the "Convolvotron", designed by Scott Foster. The device convolves an analog signal, or an optional signal file, with filter coefficients determined by the co-ordinates of the target location and the position of the listener's head, thus "placing" the signal in the perceptual 3-space of the user. The current configuration allows up to four independent and simultaneous sources and is capable of more than 300 million multiply-accumulates per second. The resulting data stream is converted to left and right analog signals and presented over headphones.

Motion trajectories and static locations at greater resolution than the empirical measurements are simulated by interpolation with linear weighting functions. When integrated with a head-tracking system, the operator's head position is monitored in realtime so that the four virtual sources are stabilized in fixed locations or in motion trajectories relative to the listener. Such head-coupling should enhance the simulation since previous studies indicate that head movements are important for localization. Informal tests at Wisconsin and at Ames also suggest that the approach is feasible; simple linear interpolations between locations as far apart as 60 degrees azimuth are perceptually indistinguishable from stimuli synthesized from measured coefficients. As with any system required to compute data "on the fly", the term "realtime" is a relative one. The digital signal-processor is designed to have a maximal latency or directional update interval of 10-30 msec, depending upon such factors as the number of simultaneous sources and the number of filter coefficients per source. Additional latencies are introduced by the headtracker (approximately 50 msec) and the IBM AT host (approximately 10-50 msec, depending upon the complexity of the source geometry). Recent work on the perception of auditory motion by Perrott and others suggests that these latencies are acceptable for moderate velocities (for 8 to 360 degrees/sec, minimum perceivable delays range from approximately 390 to 60 msec as measured by the Minimum Audible Movement Angle, 500-Hz tone-burst; Perrott & Tucker, 1988).

Psychophysical Validation.

The working assumption of the synthesis technique is that if, using headphones, one can produce ear-canal waveforms identical to those produced by a free-field source, the free-field experience would be duplicated. The only conclusive test of this assumption must come from psychophysical studies in which free-field and synthesized free-field listening are directly compared. A recent study by Wightman and Kistler (1988b) confirmed the perceptual adequacy of the basic technique for static sources; source azimuth was synthesized nearly perfectly for all listeners while source elevation was somewhat less robust and more subject to individual differences. Future research in human sound localization will continue to have a critical impact on the utility of a three-dimensional auditory display in any context. For example, an understanding of the relative contribution of listener-dependent pinna cues to localization accuracy has important implications for the ease with which such a display can be used by the general population of potential listeners. Preliminary data suggest that using non-listener-specific transforms to achieve synthesis of localized cues is at least feasible. Localization performance is only slightly degraded compared to a subject's inherent ability, even for the less robust elevation cues, as long as the transforms are derived from what one might call a "good" localizer. Further, the data suggest that individual differences in performance, particularly for elevation, can be traced to acoustical idiosyncracies in the 5 to 10kHz region of the spectrum (Wenzel, et. al., 1988b).

REFERENCES

Brooks, F.P. (1988) Grasping reality through illusion -- Interactive graphics serving science. *Proc. CHI'88, ACM Conf. Hum. Fac. Comp. Sys.*, Washington, D.C., 1-11.

Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.

Fisher, S.S., Wenzel, E.M., Coler, C. & McGreevy, M.W. (1988) Virtual interface environment workstations. *Proc. Hum. Fac. Soc.*, 32, 91-95.

Foley, J.D. (1987) Interfaces for advanced computing. *Sci. Amer.*, 257, 126-135.

Perrott, D.R. & Tucker, J. (1988) Minimum audible movement angle as a function of signal frequency and the velocity of the source. *J. Acoust. Soc. Am.*, 83, 1522-1527.

Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988a) A virtual display system for conveying three-dimensional acoustic information. *Proc. Hum. Fac. Soc.*, 32, 86-90.

Wenzel, E.M., Wightman, F.L., Kistler, D.J., & Foster, S.H. (1988b) Acoustic origins of individual differences in sound localization behavior. *J. Acoust. Soc. Amer.*, 84, S79.

Wightman, F.L. & Kistler, D.J. (1989a) Headphone simulation of free-field listening I: stimulus synthesis. *J. Acoust. Soc. Amer.*, 85, 858-867.

Wightman, F.L. & Kistler, D.J. (1989b) Headphone simulation of free-field listening II: psychophysical validation. *J. Acoust. Soc. Amer.*, 85, 868-867.

The Convolvotron®

Synthetic 3D Audio

The Convolvotron is a very high-speed digital audio signal processing system (DSP) that delivers three-dimensional sound over headphones.

The Convolvotron consists of a two-card set designed for an industry-compatible PC. The system is controlled by the host PC with calls to a library written in Microsoft C®. 128 parallel multiply/accumulate/shift processors on the two-card set make the system more than 20 times faster than ordinary DSP systems.

Programmed for a variety of signal processing tasks, such as linear filtering and time-varying filtering, the Convolvotron's primary application is the presentation of three-dimensional audio signals over headphones. In this application four independent sound sources are filtered with large, time-varying filters that compensate for the head motion of the listener and/or the possible motion of audio sources. As the listener changes the position of his or her head, the perceived location of the sound source remains constant (e.g. sound perceived to come from in front of the listener will change smoothly to the right side of the listener when he or she turns 90 degrees to the left).

Convolvotron Specifications

General:

- 2-card set for PC-AT computer. (XT version available also.)
- Inputs: 4 synchronized 16-bit Analog/Digital converters running at 50kHz.
- Outputs: 2 separate 16-bit Digital/Analog converters synchronized to A/Ds.
- 128 parallel 16 x 16 multipliers provided in each system.
- Peak convolution speed of 320 million multiply, accumulate and shifts/sec.

Auditory Display Performance:

- With 1 input source → 512 coefficients/ear, coefficients determined from a 4-way interpolation and updated at the sample rate for smooth motion. Sample rate is 50kHz. Average delay for directional controls and audio through the system is .041 seconds. 36 pairs of filters available for interpolation.
- With 2 input sources → 256 coefficients/ear, coefficients determined from a 4-way interpolation and updated at the sample rate for smooth motion. Sample rate is 50kHz. Average delay for directional controls and audio through the system is .028 seconds. 50 pairs of filters available for interpolation.
- With 4 input sources → 128 coefficients/ear, coefficients determined from a 4-way interpolation and updated at the sample rate for smooth motion. Sample rate is 50kHz. Average delay for directional controls and audio through the system is .033 seconds. 72 pairs of filters available for interpolation.

Static filters: (sample rate of 50kHz)

- With 1 input, 1 output, .358 seconds delay → 4096 coefficients.
- With 1 input, 1 output, .005 seconds delay → 1792 coefficients.
- With 1 input, 2 outputs, .04 seconds delay → 1024 coefficients/ear.

Static filters: (sample rate of 25kHz)

- With 1 input, 2 outputs, .39 seconds delay → 4096 coefficients/ear.
- With 1 input, 1 output, .006 seconds delay → 4096 coefficients.

Adaptive filter:

- LMS-type, FIR adaptive filter with 1024 coefficients, sample rate of 50kHz.

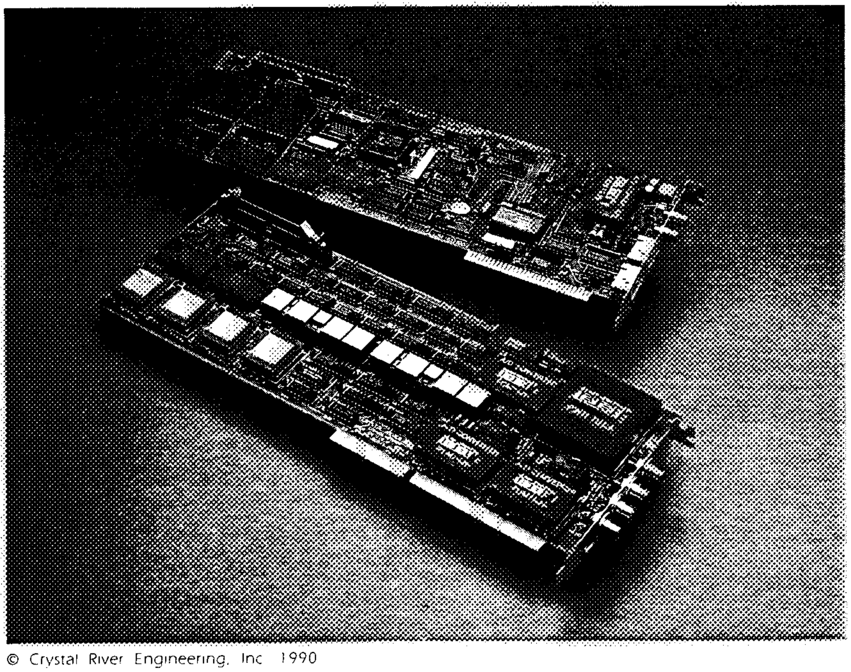
Electrical:

- Input signal range → +/- 1.0 volts. SHC5320 Sample and hold each input.
- Input impedance → 1000 Ohms.
- Output signal range → +/- 10.0 volts.
- Output impedance → 1000 Ohms.

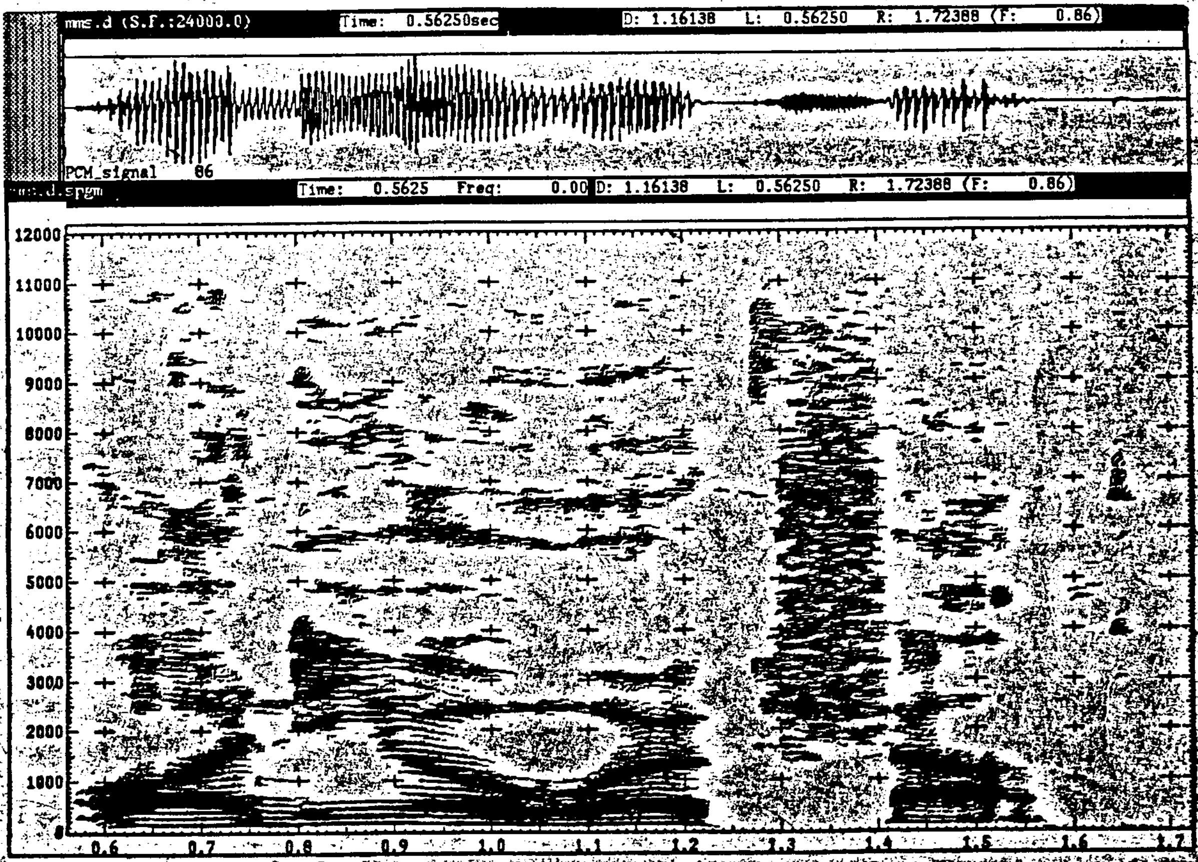
Contact:

Scott Foster
President
(209) 962-6382
Hardie Dunn
Vice President, Marketing
(209) 962-4118

Crystal River Engineering, Inc.
12350 Wards Ferry Road
Groveland, CA 95321



FINAL PROGRAM AND PAPER SUMMARIES



```
mms.d (S.F.:24000.0)   Time:  0.56250sec   D: 1.16138   L:  0.56250   R:  1.72388 (F:    0.86)
PCM_signal    86
mms.d.spgm   Time:   0.5625   Freq:      0.00 D: 1.16138   L:  0.56250   R:  1.72388 (F:    0.86)
```

1989 IEEE ASSP WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS

October 15-18, Mohonk Mountain House
New Paltz, New York

Realtime Digital Synthesis of Localized Auditory Cues over Headphones

Elizabeth M. Wenzel
Mail Stop 239-3
NASA-Ames Research Center
Moffett Field, CA 94035
beth@aurora.arc.nasa.gov
(415) 694-6290

Scott H. Foster
Crystal River Engineering
12350 Wards Ferry Road
Groveland, CA 95321
(209) 962-6382

Frederic L. Wightman, Doris J. Kistler
Waisman Center
1500 Highland Ave.
Univ. of Wisconsin
Madison, WI 53705
(608) 263-3270

Recent years have seen many advances in computing technology with the associated requirement that people be able to manage and interpret increasingly complex systems of information. As a result, an increasing amount of applied research has been devoted to reconfigurable interfaces like the virtual display. As with most research in information displays, virtual displays have generally emphasized visual information. Many investigators, however, have pointed out the importance of the auditory system as an alternative or supplementary information channel. Such a display could be realized with an array of real sound sources or loudspeakers (Doll, et. al., 1986). The signal-processing device being developed at NASA-Ames is capable of generating externalized, three-dimensional sound cues for headphone presentation in realtime (Wenzel, et. al., 1988a).

Applications of a three-dimensional auditory display involve any context in which the user's spatial awareness is important, particularly when visual cues are limited or absent. Examples include air traffic control displays for the tower or cockpit, advanced teleconferencing environments, monitoring telerobotic activities in hazardous situations, and scientific "visualization" of multi-dimensional data. (e.g., Doll, et. al., 1986; Foley, 1987; Fisher, et. al., 1988; Brooks, 1988).

Synthesizing Out-of-Head Localization.

The synthesis technique is based on the Head-Related Transfer Function (HRTF); the listener-specific, direction-dependent acoustic effects imposed on an incoming signal by the outer ears. HRTFs in the form of Finite Impulse Responses (FIRs) are measured with small probe microphones placed near the two eardrums of a listener seated in an anechoic chamber for 144 different speaker locations at intervals of 15 degrees azimuth and 18 degrees elevation (range: -36 to +54) [see Wightman & Kistler, 1989a]. In order to synthesize localized sounds, a map of listener-specific "location filters" is constructed from the 144 pairs of FIR filters represented in the time domain. The map of FIR filters is downloaded from an IBM AT to the dual-port memory of a realtime digital signal-processor, the "Convolvotron", designed by Scott Foster. The device convolves an analog signal, or an optional signal file, with filter coefficients determined by the co-ordinates of the target location and the position of the listener's head, thus "placing" the signal in the perceptual 3-space of the user. The current configuration allows up to four independent and simultaneous sources and is capable of more than 300 million multiply-accumulates per second. The resulting data stream is converted to left and right analog signals and presented over headphones.

Motion trajectories and static locations at greater resolution than the empirical measurements are simulated by interpolation with linear weighting functions. When integrated with a head-tracking system, the operator's head position is monitored in realtime so that the four virtual sources are stabilized in fixed locations or in motion trajectories relative to the listener. Such head-coupling should enhance the simulation since previous studies indicate that head movements are important for localization. Informal tests at Wisconsin and at Ames also suggest that the approach is feasible; simple linear interpolations between locations as far apart as 60 degrees azimuth are perceptually indistinguishable from stimuli synthesized from measured coefficients. As with any system required to compute data "on the fly", the term "realtime" is a relative one. The digital signal-processor is designed to have a maximal latency or directional update interval of 10-30 msec, depending upon such factors as the number of simultaneous sources and the number of filter coefficients per source. Additional latencies are introduced by the headtracker (approximately 50 msec) and the IBM AT host (approximately 10-50 msec, depending upon the complexity of the source geometry). Recent work on the perception of auditory motion by Perrott and others suggests that these latencies are acceptable for moderate velocities (for 8 to 360 degrees/sec, minimum perceivable delays range from approximately 390 to 60 msec as measured by the Minimum Audible Movement Angle, 500-Hz tone-burst; Perrott & Tucker, 1988).

Psychophysical Validation.

The working assumption of the synthesis technique is that if, using headphones, one can produce ear-canal waveforms identical to those produced by a free-field source, the free-field experience would be duplicated. The only conclusive test of this assumption must come from psychophysical studies in which free-field and synthesized free-field listening are directly compared. A recent study by Wightman and Kistler (1988b) confirmed the perceptual adequacy of the basic technique for static sources; source azimuth was synthesized nearly perfectly for all listeners while source elevation was somewhat less robust and more subject to individual differences. Future research in human sound localization will continue to have a critical impact on the utility of a three-dimensional auditory display in any context. For example, an understanding of the relative contribution of listener-dependent pinna cues to localization accuracy has important implications for the ease with which such a display can be used by the general population of potential listeners. Preliminary data suggest that using non-listener-specific transforms to achieve synthesis of localized cues is at least feasible. Localization performance is only slightly degraded compared to a subject's inherent ability, even for the less robust elevation cues, as long as the transforms are derived from what one might call a "good" localizer. Further, the data suggest that individual differences in performance, particularly for elevation, can be traced to acoustical idiosyncracies in the 5 to 10kHz region of the spectrum (Wenzel, et. al., 1988b).

The realtime system will be available for demonstration to interested listeners at the Workshop.

REFERENCES

Brooks, F.P. (1988) Grasping reality through illusion – Interactive graphics serving science. *Proc. CHI'88, ACM Conf. Hum. Fac. Comp. Sys.*, Washington, D.C., 1-11.

Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.

Fisher, S.S., Wenzel, E.M., Coler, C. & McGreevy, M.W. (1988) Virtual interface environment workstations. *Proc. Hum. Fac. Soc.*, 32, 91-95.

Foley, J.D. (1987) Interfaces for advanced computing. *Sci. Amer.*, 257, 126-135.

Perrott, D.R. & Tucker, J. (1988) Minimum audible movement angle as a function of signal frequency and the velocity of the source. *J. Acoust. Soc. Am.*, 83, 1522-1527.

Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988a) A virtual display system for conveying three-dimensional acoustic information. *Proc. Hum. Fac. Soc.*, 32, 86-90.

Wenzel, E.M., Wightman, F.L., Kistler, D.J., & Foster, S.H. (1988b) Acoustic origins of individual differences in sound localization behavior. *J. Acoust. Soc. Amer.*, 84, S79.

Wightman, F.L. & Kistler, D.J. (1989a) Headphone simulation of free-field listening I: stimulus synthesis. *J. Acoust. Soc. Amer.*, 85, 858-867.

Wightman, F.L. & Kistler, D.J. (1989b) Headphone simulation of free-field listening II: psychophysical validation. *J. Acoust. Soc. Amer.*, 85, 868-867.

Session AA. Physiological Acoustics VI and Psychological Acoustics III: Localization and Binaural Hearing (Poster Session)

Whitlow W. L. Au, Cochairman
Naval Ocean Systems Center
P. O. Box 997
Kailua, Hawaii 96734

Masanao Ebata, Cochairman
Faculty of Engineering
Kumamoto University
2-39-1 Kurokami
Kumamoto, 860 Japan

Contributed Papers

All posters will be displayed from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of papers AA1 through AA7 will be at their posters from 8:30 to 10:00 a.m. contributors of papers AA8 through AA13 will be at their posters from 10:00 to 11:30 a.m.

AA1. The Franssen effect and the localization plausibility hypothesis. Brad Rakerd (Department of Audiology and Speech Sciences, Michigan State University, East Lansing, MI 48824) and William Morris Hartmann (Department of Physics, Michigan State University, East Lansing, MI 48824)

The Franssen effect is obtained with two loudspeakers in a room. If a sine tone is abruptly turned on at the left loudspeaker, then slowly faded off while the right loudspeaker is slowly faded on, a listener will judge that the tone continues to come from the left loudspeaker, even though the left loudspeaker is not sounding at all. Our own studies of localization of sound in rooms have led to a principle of localization called the "plausibility hypothesis." One of the predictions of this hypothesis is that in an anechoic room the Franssen effect should fail [B. Rakerd and W. M. Hartmann, J. Acoust. Soc. Am. 78, 524–533 (1985)]. Experimental studies are reported using the Franssen stimulus both in an ordinary room and in an anechoic room. The results of the experiment support the prediction. [Work supported by the National Institutes of Health.]

AA2. Acoustic origins of individual differences in sound localization behavior. Elizabeth Wenzel, Frederic Wightman, Doris Kistler, and Scott Foster (NASA-Ames Research Center, Moffett Field, CA 94035, Department of Psychology and Waisman Center, University of Wisconsin, Madison, WI 53705, and Crystal River Engineering, 12350 Wards Ferry Road, Groveland, CA 95321)

Human listeners vary widely in their ability to localize unfamiliar sounds in an environment devoid of visual cues. Our research, in which blindfolded listeners give numerical estimates of apparent source azimuth and elevation, suggests that individual differences are greatest in judgments of source elevation: listeners are uniformly accurate when judging source azimuth. The pattern of individual differences is the same for free-field sources and for simulated free-field sources presented over headphones. Simulated free-field sources are produced by digital filtering techniques which incorporate the listener-specific, direction-dependent acoustic effects of the outer ears. Two features of this data bear on the question of the origin of individual differences in elevation accuracy: (1) a listener's accuracy in judging source elevation can be predicted from an analysis of the acoustic characteristics of the listener's outer ears; (2) the pattern of elevation errors made by one listener (A) can be transferred to another listener (B) by presenting to listener B the simulated free-field sources derived from the outer-ear acoustics of listener A. Thus it is believed that many of the individual differences in localization behavior are traceable to individual differences in outer-ear acoustics. The data have important implications for the study of localization in both basic and applied contexts. [Work supported by NASA, NSF, and USAF-AAMRL-AFSC.]

AA3. Prefiltering method for a head-related stereophonic system. Takayuki Mizuuchi, Kaoru Okabe, Hareo Hamada, and Tanetoshi Miura (Tokyo Denki University, 2-2 Kanda-Nishiki-cho, Chiyoda-ku, Tokyo, 101 Japan)

The prefiltering scheme of the usual head-related stereophonic system using two loudspeakers for reproduction [M. R. Schroeder *et al.*, J. Acoust. Soc. Am. 56, 1195 (1974)] was modified. In addition to the usual filtering scheme, an additional filtering stage is introduced that converted the frontal incident characteristics of the dummy head into those for individual listeners. Using this filtering scheme, the exact reproduction of the frontal sound image becomes possible, which is difficult when the sound recorded through a dummy head is played back. The ability of this modified system to reproduce the original sound localization by a listening test in an anechoic chamber is evaluated. In the test, both the conventional scheme and the new one were evaluated. As a result, the localization of the frontal sound image was reproduced exactly using the new system, and the conventional system failed in this reproduction. It was also confirmed that sound from other directions in the three-dimensional space was almost perfectly reproduced using the new system.

AA4. Echo suppression or localization masking? Pierre L. Divenyi (Speech and Hearing Research, V. A. Medical Center, Martinez, CA 94553)

A brief dichotic conditioner (C) has been shown to effectively disrupt lateralization of a brief probe (P) presented after a short interval (4–7 ms, onset to onset) with an interaural time delay that is perfectly audible in the absence of the C [P. M. Zurek, J. Acoust. Soc. Am. 66, 1750–1757 (1980)], even when the C and the P are different sounds [P. L. Divenyi and J. Blauert, in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ, 1987), pp. 146–155]. An echo suppression mechanism responsible for this effect would predict (1) suppression to be strongest when the C and the P are identical and to decrease monotonically as the spectra of the two sounds are made different, and (2) a monotonic falloff of suppression when the temporal separation between the two sounds is increased beyond a certain minimum. In the present experiments, the effects of frequency separation and temporal separation between a narrow-band P centered at 2 kHz and a C were systematically explored. Contrary to the predictions, low-frequency ($0.8 < 1.3$ kHz) C's were more effective in suppressing the lateralization of the P than those closer to the P frequency, and lateralization performance was nonmonotonically related to temporal separation between C and P. The results suggest that a dichotic stimulus with a relatively high *localization strength* could "mask" the localization of another, subsequent dichotic stimulus. [Work supported by the Veterans Administration.]

Conference handout for Abs[illegible]

Conference handout

The Convolvotron: Realtime Synthesis of Out-of-Head Localization

Joint Meeting of the Acoustical Society of America and
the Acoustical Society of Japan
Honolulu, HI, November 14-18, 1988

Elizabeth M. Wenzel
Aerospace Human Factors Research Division
NASA-Ames Research Center
Mail Stop 239-3
Moffett Field, CA 94035

Frederic L. Wightman, Doris J. Kistler
Waisman Center
University of Wisconsin
Madison, WI 53705

Scott H. Foster
Crystal River Engineering
12350 Wards Ferry Road
Groveland, CA 95321

The Convolvotron: Realtime Synthesis of Out-of-Head Localization

Elizabeth M. Wenzel
NASA-Ames Research Center
Moffett Field, CA

Frederic L. Wightman, Doris J. Kistler
Univ. of Wisconsin
Madison, WI

Scott H. Foster
Crystal River Engineering
Groveland, CA

Recent years have seen many advances in computing technology with the associated requirement that people be able to manage and interpret increasingly complex systems of information. As a result, an increasing amount of applied research has been devoted to reconfigurable interfaces like the virtual display. As the technology has advanced, virtual displays have gone beyond the flat screen, assuming a three-dimensional spatial organization which provides a more natural means of accessing and manipulating information. A few projects, such as the Virtual Interactive Environment Workstation (VIEW) at NASA's Ames Research Center, have taken the spatial metaphor to its limit by directly involving the operator in the data environment (Fisher, et. al., 1986; 1988). The kind of "artificial reality" once relegated solely to the specialized world of the cockpit is now being seen as the next step in interface development for all types of advanced computing applications (Foley, 1987). The goal is to create a highly flexible and interactive simulation which integrates visual, auditory, tactile, and kinesthetic cues into a complex three-dimensional, virtual world.

As with most research in reconfigurable information displays, virtual displays have generally emphasized visual information. Many investigators, however, have pointed out the importance of the auditory system as an alternative or supplementary information channel. We believe that a three-dimensional auditory display can enhance information transfer by combining positional and semantic information to form dynamic, multidimensional acoustic events which convey meaning about objects in the spatial world of the listener. Such a display could be realized with an array of real sound sources or loudspeakers (Doll, et. al., 1985). At NASA's Ames Research Center, we're developing a signal-processing device which maximizes flexibility and portability by synthetically generating three-dimensional sound in realtime for delivery through headphones (Wenzel, et. al., 1988).

Synthesizing Out-of-Head Localization.

The synthesis technique is based on the Head-Related Transfer Function (HRTF); the listener-specific, direction-dependent acoustic effects imposed on an incoming signal by the outer ears. HRTFs in the form of Finite Impulse Responses are measured with small probe microphones placed near the two eardrums of a listener seated in an anechoic chamber for 144 different speaker locations at intervals of 15 degrees azimuth and 18 degrees elevation (range: -36 to +54) [see Wightman & Kistler, in press].

In order to synthesize localized sounds, a map of "location filters" is constructed from the 144 pairs of FIR filters by transforming to the frequency domain, removing the spectral effects of the original loudspeakers and headphones using Fourier techniques, and then transforming back to the time domain. The map of corrected FIR filters is downloaded from an IBM AT to the dual-port memory of a realtime digital signal-processor designed by Scott Foster and currently being prototyped as the "The Convolvotron" (Figure 1). The device convolves an analog signal, or an optional signal file, with filter coefficients determined by the co-ordinates of the target location and the position of the listener's head, thus "placing" the signal in the perceptual 3-space of the user. The initial configuration allows up to three independent and simultaneous sources and is capable of more than 300 million multiply-accumulates per second. The resulting data stream is converted to left and right analog signals and presented over headphones.

Motion trajectories and static locations at greater resolution than the empirical measurements are simulated by interpolation with linear weighting functions. When integrated with the head-tracking system, the operator's head position is monitored in realtime so that the three simultaneous sources are stabilized in fixed locations or in motion trajectories relative to the listener. Such head-coupling should enhance the simulation since previous studies indicate that head movements are important for localization (Thurlow & Runge, 1967). Informal tests at Wisconsin and at Ames also suggest that the approach is feasible; simple linear interpolations between locations as far apart as 60 degrees azimuth are perceptually indistinguishable from stimuli synthesized from measured coefficients. As with any system required to compute data "on the fly", the term "realtime" is a relative one. The digital signal-processor is designed to have a maximal latency or directional update interval of 10-30 msec, depending upon such factors as the number of simultaneous sources and the number of filter coefficients per source. Additional latencies are introduced by the headtracker (approximately 50 msec) and the IBM AT host (approximately 10-50 msec, depending upon the complexity of the source geometry). Recent work on the perception of auditory motion by Perrott and others suggests that these latencies are acceptable for moderate velocities (for 8 to 360 degrees/sec, minimum perceivable delays range from approximately 390 to 60 msec as measured by the Minimum Audible Movement Angle, 500-Hz tone-burst; Perrott, 1982; Perrott & Tucker, 1988).

Applications.

Three-dimensional auditory displays will be most usefully applied in contexts where the user's representation of spatial information is important. Direct knowledge of spatial relationships is particularly critical under conditions of high cognitive and motor workload, especially when visual cues are degraded or absent and immediate sensation of the auditory world may not be possible or desirable. Examples of such tasks include monitoring one's location relative to other objects or vehicles during Extra-Vehicular Activity in space, air traffic control displays for the tower or cockpit, supervisory control of telerobotic/telepresence activities, scientific "visualization" of multi-dimensional data as in computational fluid dynamics and molecular modelling, acoustic navigation displays for the blind, and advanced teleconferencing environments (e.g., see Doll, et. al., 1986; Fisher, et. al., 1986, 1988; Wenzel, et. al., 1988; Brooks, 1988; Hart, 1988).

Research devoted to understanding basic mechanisms of human sound localization will have a critical impact on the utility of a three-dimensional auditory display in any context. For example, an understanding of the relative contribution of listener-dependent pinna cues to localization accuracy has important implications for the ease with which such a display can be used by the general population of potential listeners. The preliminary data suggest that using non-listener-specific transforms to achieve synthesis of localized cues is at least feasible. Localization performance is not degraded much beyond a subject's inherent ability, even for the less robust elevation cues, as long as the transforms are derived from what one might call a "good" localizer.

It should also be remembered that constraints may be imposed by the actual application of knowledge derived from basic research. The goal of a 3D auditory display is to present unambiguous spatial information as flexibly, dynamically, and efficiently as possible, often under conditions which are less than ideal for detecting subtle acoustic cues. Ultimately, many issues will need to be investigated and many compromises made in attempting to produce veridical acoustic displays. Relatively little is known about the nature of auditory motion perception (Perrott, 1982; Grantham, 1986), the role of auditory psychomotor co-ordination (Perrott, et. al., 1988), the critical cues for distance (Coleman, 1962), or the potential contribution of reverberation (Mershon & King, 1975) to an accurate representation of auditory space. Earlier work also suggests that out-of-head localization is influenced by the nature of accompanying visual cues (Gardner, 1968; Mershon, et. al., 1980), and the listener's familiarity with the sounds to be localized (Coleman, 1962). In the past it has not been possible to fully explore many aspects of localization simply because it was technically too difficult to put the stimuli under direct experimental control. We hope that realtime signal-processing devices like the prototype being developed at Ames will prove to be useful tools for examining some of these issues as well as furnish the basic technology for sophisticated acoustic displays.

REFERENCES

Brooks, F.P. (1988) Grasping reality through illusion – Interactive graphics serving science. *Proc. CHI'88, ACM Conf. Hum. Fac. Comp. Sys.*, Washington, D.C., 1-11.

Coleman, P.D. (1962) Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. Am.*, 34, 345-346.

Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.

Fisher, S.S., Mc Greevy, M., Humphries, J., & Robinett, W. (1986) Virtual Environment Display System. *ACM Workshop on Interactive 3D Graphics*, Oct. 23-24, Chapel Hill, NC.

Fisher, S.S., Wenzel, E.M., Coler, C. & McGreevy, M.W. (1988) Virtual interface environment workstations. *Proc. Hum. Fac. Soc.*, 32, 91-95.

Foley, J.D. (1987) Interfaces for advanced computing. *Sci. Amer.*, 257, 126-135.

Gardner, M.B. (1968) Proximity image effect in sound localization. *J. Acoust. Soc. Am.*, 43, 163.

Grantham, D.W. (1986) Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J. Acoust. Soc. Am.*, 79, 1939-1949.

Hart, S.G. (1988) Helicopter human factors. In *Human Factors in Aviation*, E. Wiener & D. Nagel (Eds.), Academic Press: New York.

Mershon, D.H. & King, L.E. (1975) Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perc. & Psychophys.*, 18, 409-415.

Mershon, D.H., Desaulniers, D.H., Amerson, Jr., T.L., & Kiefer, S.A. (1980) Visual capture in auditory distance perception: Proximity image effect reconsidered. *J. Aud. Res.*, 20, 129-136.

Perrott, D.R (1982) Studies in the perception of auditory motion. In *Localization of Sound: Theory and Applications*, R.W. Gatehouse (Ed.), Amphora Press: Groton, CN, 169-193.

Perrott, D.R., Ambarsoom, H., & Tucker, J. (1987) Changes in head position as a measure of auditory localization performance: Auditory psychomotor coordination under monaural and binaural listening conditions. *J. Acoust. Soc. Am.*, 82, 1637-1645.

Perrott, D.R. & Tucker, J. (1988) Minimum audible movement angle as a function of signal frequency and the velocity of the source. *J. Acoust. Soc. Am.*, 83, 1522-1527.

Thurlow, W.R. & Runge, P.S. (1967) Effect of induced head movements on localization of direction of sounds. *J. Acoust. Soc. Am.*, 42, 480-488.

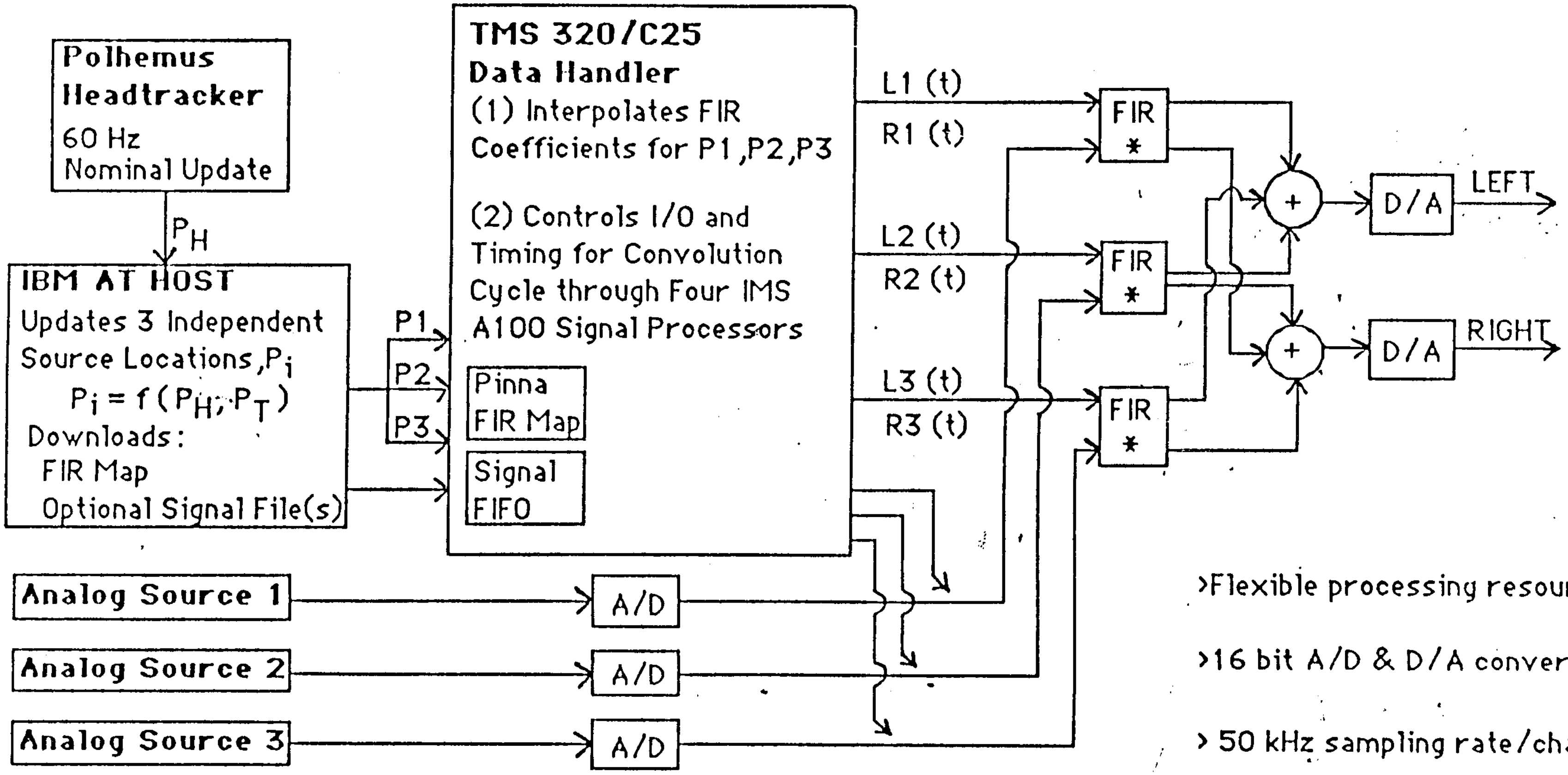
Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988) A virtual display system for conveying three-dimensional acoustic information. *Proc. Hum. Fac. Soc.*, 32, 86-90.

Wightman, F.L. & Kistler, D.J. (1989a) Headphone simulation of free-field listening I: stimulus synthesis. *J. Acoust. Soc. Amer.*, (In press).

Wightman, F.L. & Kistler, D.J. (1989b) Headphone simulation of free-field listening II: psychophysical validation. *J. Acoust. Soc. Amer.*, (In press).

THE CONVOLVOTRON

High-speed realtime digital signal-processor



Polhemus Headtracker
60 Hz
Nominal Update

P_H

IBM AT HOST
Updates 3 Independent Source Locations, P_i
$P_i = f(P_H; P_T)$
Downloads:
FIR Map
Optional Signal File(s)

P1
P2
P3

TMS 320/C25
Data Handler
(1) Interpolates FIR Coefficients for P1, P2, P3

(2) Controls I/O and Timing for Convolution Cycle through Four IMS A100 Signal Processors

Pinna FIR Map

Signal FIFO

L1 (t)
R1 (t)
L2 (t)
R2 (t)
L3 (t)
R3 (t)

FIR *
FIR *
FIR *

+
+

D/A
D/A

LEFT
RIGHT

Analog Source 1
Analog Source 2
Analog Source 3

A/D
A/D
A/D

>Flexible processing resources

>16 bit A/D & D/A conversion

> 50 kHz sampling rate/channel

> Processing rate approx. 300 million multiply-accumulates per sec (4 IMS A100 chips)

> Estimated latencies:
Headtracker; 50 msec
IBM AT; 10-50 msec
Signal-processor; 10-30 msec

1988 Conference Handout - Acoustical Society of America, 84, S79

Acoustical Origins of Individual Differences in Sound Localization Behavior

Acoustical and psychophysical techniques are used to study individual differences in sound localization behavior. The free-field-to-eardrum transfer function of each subject is measured. Probe microphones are placed approximately 1 mm from the eardrums in a listener's unoccluded ear canals. Using a periodic wideband source, the acoustical transfer functions of the listener's two ears are measured for sound sources at 144 positions all around the listener, at elevations from -36° to +54°. These transfer functions are highly variable from listener to listener. Figure 1 shows the mean standard deviation of 1/3-octave band levels across the 10 listeners and 144 source positions. The range of the 1/3-octave band levels is indicated by the dashed lines. Note that in the frequency range from 5 kHz to 10 kHz, the individual differences are greatest.

Free-field sound sources are simulated by presenting digitally processed stimuli over headphones. Stimuli are passed through digital filters which incorporate the listener-specific free-field-to-eardrum transfer functions, as measured as above. The resulting stimuli, presented over headphones, produce the same waveforms at a listener's eardrums as real free-field sources. The perceptual adequacy of this simulation is assessed by comparing listeners' judgements of the apparent positions of free-field stimuli to their judgements of the apparent positions of headphone-presented stimuli. Figure 2 shows data from 2 subjects (SDO and SDE) in the form of scatterplots of perceived source azimuth and elevation vs. actual source azimuth and elevation for 72 different source positions. The close correspondence between the data obtained with free-field sources (left panels) and that obtained using simulated sources (right panels) is evident in the data from each of our 11 subjects. Table 1 gives the correlations between perceived and actual azimuth (and elevation) for both free-field and headphone conditions, for all 11 subjects. Note that substantial individual differences appear only in the elevation component of the data.

Using simulated free-field sources, idiosyncrasies of elevation perception can be transferred from one listener to another.

One's ability to localize, and especially to determine the elevation of a sound source, appears to depend on the use of acoustical cues provided by one's own ears. Electronically interchanging ears with someone else degrades perception of elevation, especially if the cues provided by the other person's ears are impoverished. Figures 3-5 show the results of three experiments in which subjects listened through other subjects' ears. The leftmost and rightmost panels in each figure show scatterplots of judgements (azimuth and elevation separately) made by subjects listening through their own ears. The middle panels show data obtained when the subject from the left panel listened through the ears of the subject from the right panel.

If a subject with "good" ears (SDP) listens through the ears of a subject with "bad" ears, (SDE) there is a marked degradation of ability to determine source elevation (see Figure 3). The data shown in Figure 4 suggests that the elevation perception of a subject with "bad" ears (SDE) is not improved by listening through the ears of a subject with "good" ears

(SDO). There are slight degradations of both azimuth and elevation perception whenever one listens through another person's ears, even if both individuals (e.g., SDP and SDO) appear to have "good" ears (see Figure 5).

Acoustical evidence on the importance of the 5-10 kHz region

Although there are large inter-subject differences in the free-field-to-eardrum transfer functions, the way those functions change with source elevation is remarkably constant across listeners, with the largest changes evident in the 5-10 kHz region. "Interaural elevation dependency" functions show how the free-field-to-eardrum transfer functions change with elevation. These functions are computed by dividing all leading ear functions by the corresponding trailing ear functions to produce "interaural difference" functions. Next, all the functions from a given azimuth are divided by the function at the same azimuth and zero elevation. These normalized elevation-dependency functions are then averaged over all azimuths. The result is a single set of functions which shows the change in interaural intensity difference caused by moving a source from zero elevation to five other elevations. Figure 6 shows elevational dependency functions from 4 subjects, including 3 subjects with "good" ears, and the 1 subject (SDE) with "bad" ears.

Psychophysical evidence on the importance of the 5-10 kHz region.

Acoustical cues in the 5-kHz to 10-kHz frequency range appear to determine perceived sound source elevation. High-pass filtering the stimuli at 5 kHz has little effect, but high-pass filtering at 10 kHz nearly eliminates elevation perception. Although somewhat affected, azimuth perception appears to be less degraded by these manipulations. Figure 7 shows data from one subject (SDO) who estimated the apparent positions of wideband (200 Hz to 14 kHz) stimuli (left panel), stimuli high-pass filtered at 5 kHz (middle panel), and stimuli high-pass filtered at 10 kHz (right panel.)

Table 1

Subject ID	Azimuth Correlation	Elevation Correlation
SDE	**.98** (.97)	**.68** (.43)
SDH	**.96** (.95)	**.92** (.83)
SDL	**.98** (.97)	**.89** (.85)
SDM	**.98** (.99)	**.94** (.93)
SDO	**.99** (.99)	**.94** (.92)
SDP	**.99** (.99)	**.96** (.88)
SED	**.97** (.98)	**.93** (.82)
SER	**.99** (.99)	**.96** (.94)
SET	**.98** (.99)	**.93** (.87)
SGB	**.99** (.97)	**.95** (.93)
SGE	**.98** (.98)	**.94** (.86)

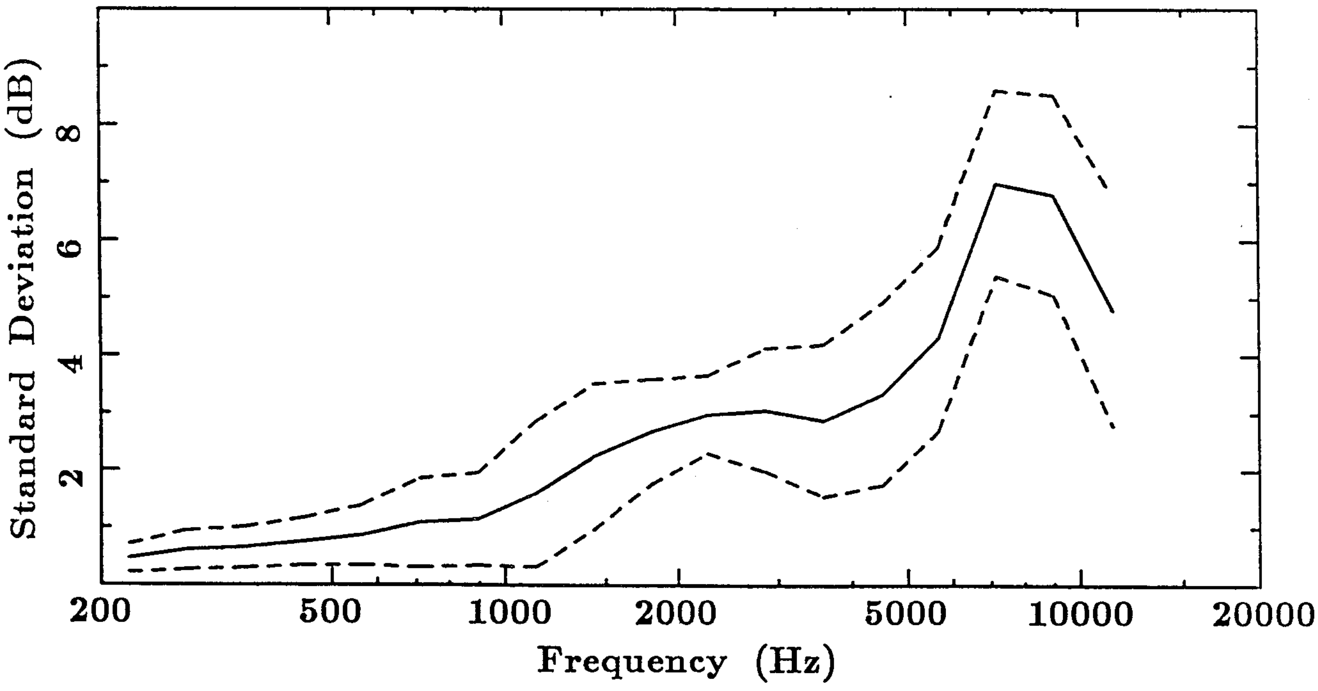

Standard Deviation (dB)
2
4
6
8
200
500
1000
2000
5000
10000
20000
Frequency (Hz)

Figure 1

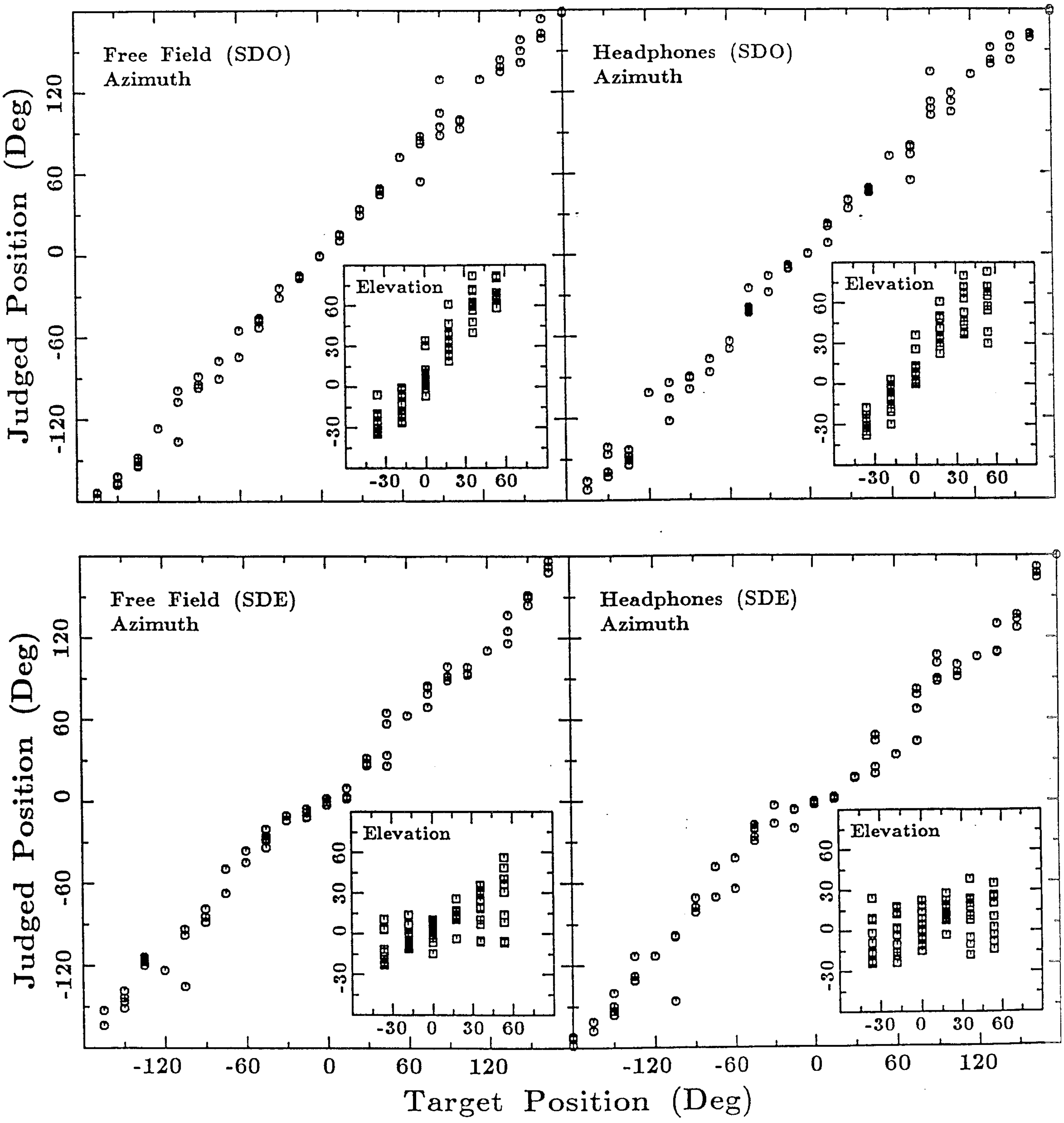

Free Field (SDO)
Azimuth
Headphones (SDO)
Azimuth
Free Field (SDE)
Azimuth
Headphones (SDE)
Azimuth
Elevation
-30
0
30
60
120
-60
-120
Judged Position (Deg)
Target Position (Deg)

Figure 2

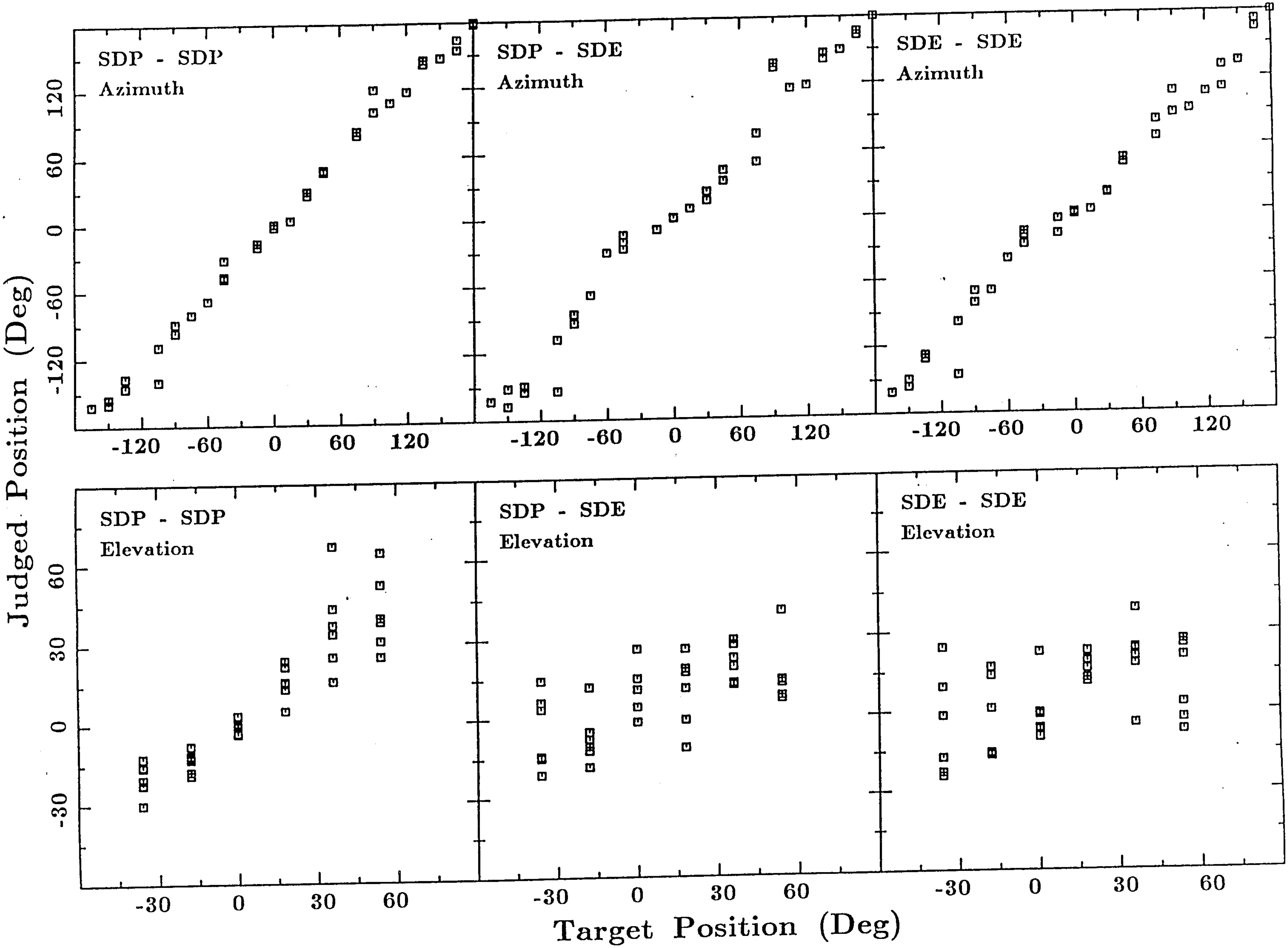

SDP - SDP
Azimuth
SDP - SDE
Azimuth
SDE - SDE
Azimuth
SDP - SDP
Elevation
SDP - SDE
Elevation
SDE - SDE
Elevation
Judged Position (Deg)
Target Position (Deg)
-120
-60
0
60
120
-30
0
30
60

Figure 3

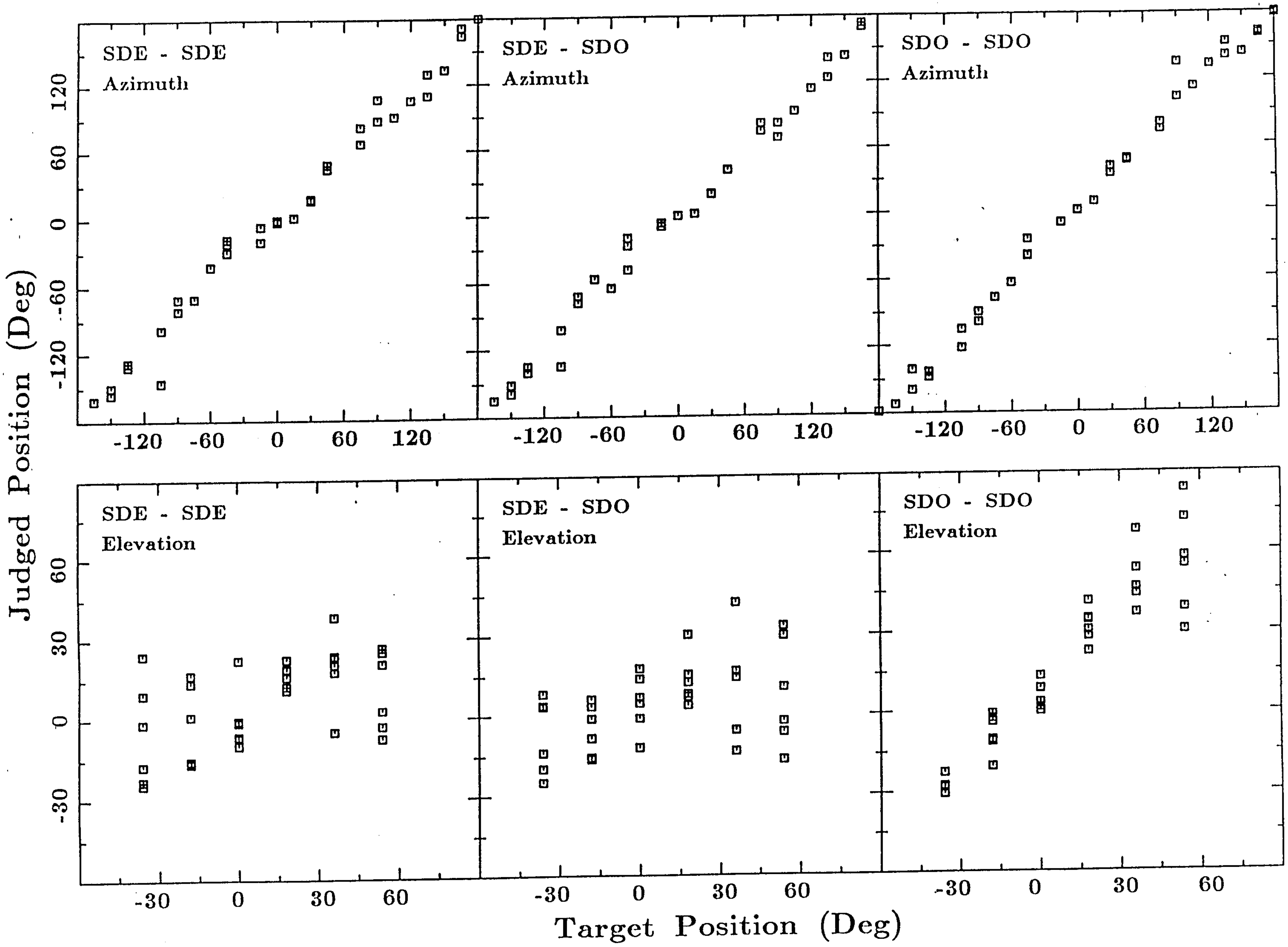

SDE - SDE
Azimuth
SDE - SDO
Azimuth
SDO - SDO
Azimuth
SDE - SDE
Elevation
SDE - SDO
Elevation
SDO - SDO
Elevation
120
60
0
-60
-120
-120
-60
0
60
120
60
30
0
-30
-30
0
30
60
Judged Position (Deg)
Target Position (Deg)

Figure 4

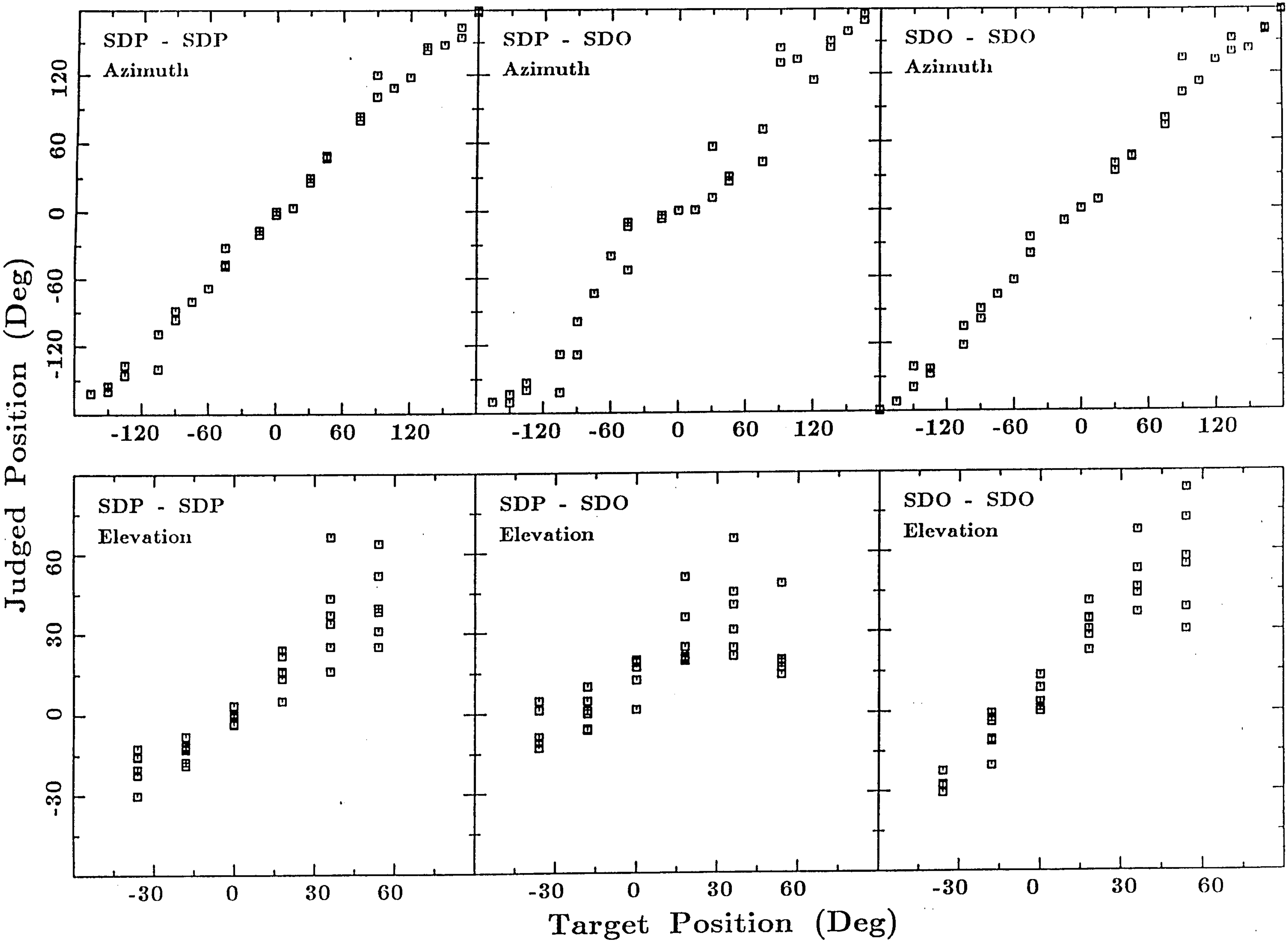

SDP - SDP
Azimuth
SDP - SDO
Azimuth
SDO - SDO
Azimuth
SDP - SDP
Elevation
SDP - SDO
Elevation
SDO - SDO
Elevation
Judged Position (Deg)
120
60
0
-60
-120
60
30
0
-30
Target Position (Deg)
-120
-60
0
60
120
-30
0
30
60

Figure 5

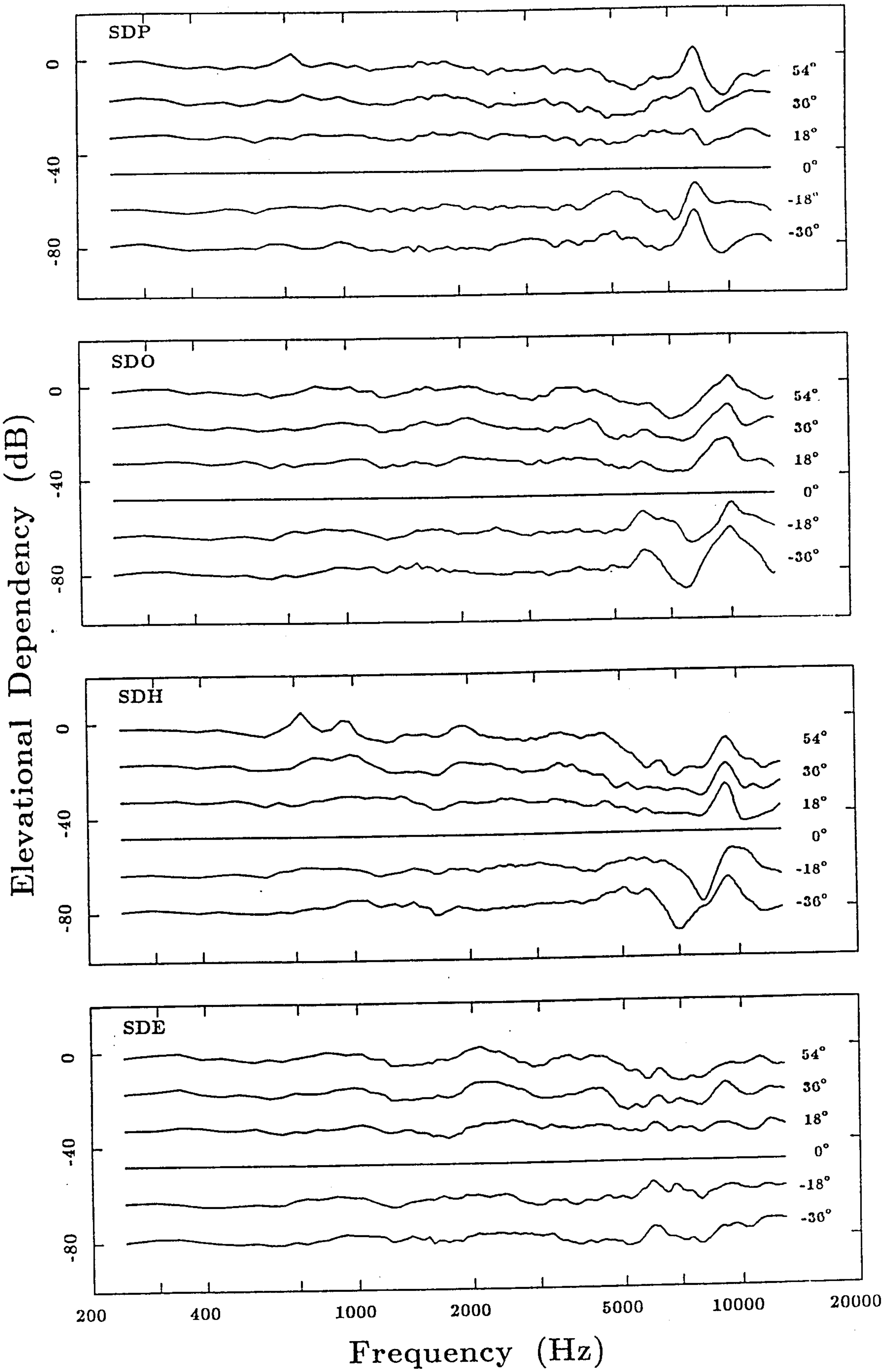

SDP
0
-40
-80
54°
36°
18°
0°
-18°
-36°
SDO
0
-40
-80
54°
36°
18°
0°
-18°
-36°
SDH
0
-40
-80
54°
36°
18°
0°
-18°
-36°
SDE
0
-40
-80
54°
36°
18°
0°
-18°
-36°
Elevational Dependency (dB)
200
400
1000
2000
5000
10000
20000
Frequency (Hz)

Figure 6

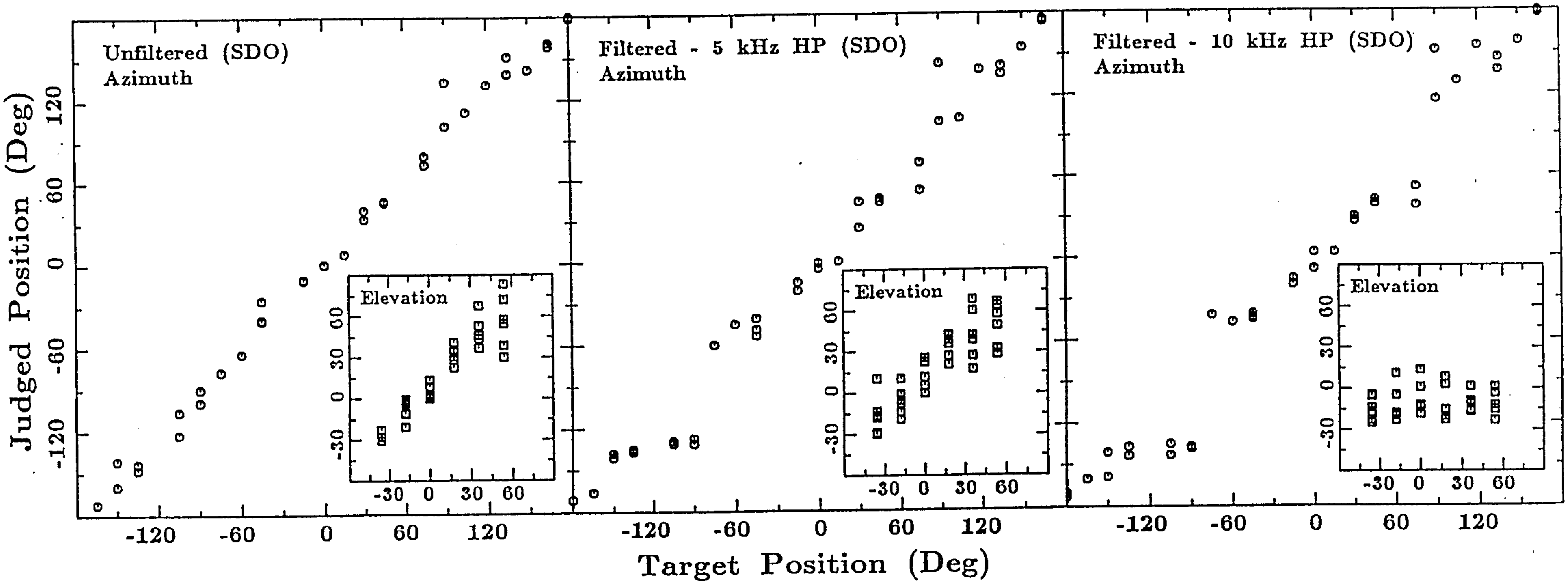

Unfiltered (SDO)
Azimuth
Filtered - 5 kHz HP (SDO)
Azimuth
Filtered - 10 kHz HP (SDO)
Azimuth
Elevation
Judged Position (Deg)
Target Position (Deg)
-120
-60
0
60
120
-30
30
60

Figure 7

A VIRTUAL DISPLAY SYSTEM FOR CONVEYING THREE-DIMENSIONAL ACOUSTIC INFORMATION

Elizabeth M. Wenzel
NASA-Ames Research Center
Moffett Field, CA

Frederic L. Wightman
University of Wisconsin
Madison, WI

Scott H. Foster
Crystal River Engineering
Groveland, CA

ABSTRACT

A three-dimensional auditory display could take advantage of intrinsic sensory abilities like localization and perceptual organization by generating dynamic, multidimensional patterns of acoustic events that convey meaning about objects in the spatial world. Applications involve any context in which the user's situational awareness is critical, particularly when visual cues are limited or absent; e.g., air traffic control or telerobotic activities in hazardous environments. Such a display would generate localized cues in a flexible and dynamic manner. Whereas this can be readily achieved with an array of real sound sources or loudspeakers, the NASA-Ames prototype maximizes flexibility and portability by synthetically generating three-dimensional sound in realtime for delivery through headphones. Psychoacoustic research suggests that perceptually-veridical localization over headphones is possible if both the direction-dependent pinna cues and the more well understood cues of interaural time and intensity are adequately synthesized. Although the realtime device is not yet complete, recent studies at the University of Wisconsin have confirmed the perceptual adequacy of the basic approach to synthesis.

INTRODUCTION

With rapid advances in technology and the concomitant requirement for managing complex informational systems, an increasing amount of research has been devoted to reconfigurable interfaces like the virtual display. As computer graphics become more sophisticated, virtual displays are assuming a three-dimensional spatial organization, providing a more natural means of interaction which may improve access and manipulation of data. Recent projects, building upon Sutherland's work (1968) on binocular head-mounted displays, have taken the spatial metaphor the farthest by directly involving the operator in the data environment. The goal is to create a highly flexible and interactive simulation/telepresence environment which integrates visual, auditory, tactile, and kinesthetic cues into a complex three-dimensional, virtual world (Fisher, this volume; Furness, 1986; Brooks, 1988).

As with most research in information displays, virtual displays have generally emphasized visual information. Many investigators, however, have pointed out the importance of the auditory system as an information channel. We believe that a three-dimensional auditory display can substantially enhance situational awareness by combining spatial and semantic information to form dynamic, multidimensional patterns of acoustic events which convey meaning about objects in the spatial world of the user. Such a display can be realized with an array of real sound sources or loudspeakers (Doll et. al., 1986). The signal-processing device being developed at NASA-Ames maximizes flexibility and portability by synthetically generating three-dimensional sound in realtime for delivery through headphones. Unlike conventional stereo, sources can be perceived outside the head at discrete distances and directions from the listener. The 3-D auditory display will be integrated with Ames' Virtual Interactive Environment Workstation (VIEW) which allows the user to explore and interact with a 360-degree synthesized or remotely-sensed world using a head-mounted, wide-angle, stereoscopic display controlled by operator position, voice, and gesture (Fisher, this volume).

SYNTHESIS OF LOCALIZED SOUND

Theory

The "Duplex Theory" of human sound localization, first described by Lord Rayleigh (1907), states that the two cues primarily responsible for localization are interaural differences in time of arrival for low-frequency waveforms and interaural differences in intensity at high frequencies. Although the theory has long dominated thinking about binaural hearing, recent research points to serious limitations with this hypothesis. For example, it cannot account for the ability of subjects to precisely locate sounds on the median plane where interaural cues are minimal or absent (see Blauert, 1983). Similarly, when subjects listen to stimuli over headphones, they are perceived as inside the head even though interaural temporal and intensity differences appropriate to an external source location are introduced. A variety of research indicates that deficiencies of the Duplex Theory stem from an inadequate consideration of filtering by the outer ears or pinnae; Shaw (1974) demonstrated that spectral shaping by the pinnae is indeed direction-dependent, others showed that the absence of pinna cues degrades localization accuracy (Gardner & Gardner, 1973; Oldfield & Parker, 1984), and Plenge (1974) concluded that pinna cues are primarily responsible for externalization or the sensation of "out-there-ness". Such studies suggest that perceptually-veridical localization over headphones is possible if both the direction-dependent pinna cues and the more well understood cues of interaural time and intensity can be adequately synthesized.

Technique

A number of techniques for synthesizing various features of auditory spatial perception have been explored. Among these are simulation of elevation cues with comb-filtering techniques, creation of auditory "spaciousness" using reverberation effects, methods of binaural recording, structural modelling of the pinnae, and the measurement of pinna transform functions (see Blauert, 1983 for discussions of these topics). Our technique involves preserving the frequency-dependent interaural time (phase) and intensity information with precise measurements of Head-Related Transfer Functions (HRTFs); the listener-specific, direction-dependent acoustic effects imposed on an incoming signal by the outer ears.

Probe microphones are placed near the two eardrums of a listener seated in the anechoic chamber of the Psychoacoustics Lab at Wisconsin. The subject is at the center of a semicircular arc (1.4 m radius) with eight loudspeakers mounted at 18 degree intervals of elevation (range: -36 to 90 degrees where 0 is at ear level). The arc is then re-positioned at 15-degree increments of azimuth (range: 0 to 360 degrees) in order to sample most of the spherical space. The stimulus, a 50-Hz train of 20.48-msec bursts of pseudorandom noise, is presented over one of the speakers at a level of approximately 70 dB SPL and the responses of the probe microphones to 1000 repetitions of the noise-burst are averaged (digitized with 16-bit A/D converters at a sampling rate of 50 kHz). A new pair of HRTFs or pinna transforms is then measured for each location.

Although the HRTFs are measured in the time domain, application of the Fourier Transform and linear systems theory allows analysis of the measurements in the frequency domain and the construction of pairs of digital filters for each location. Thus, the eight panels in Figure 1 exemplify the data with plots of the amplitude (relative decibels) and phase (radians) responses derived from the Fourier Transforms of the empirical measurements at the left (solid lines) and right ears (dotted lines) of a single listener. The HRTFs shown are for source locations at 0 degrees elevation and 0, 45, 135, or 270 degrees azimuth (0 is directly in front, degrees increase counter-clockwise).

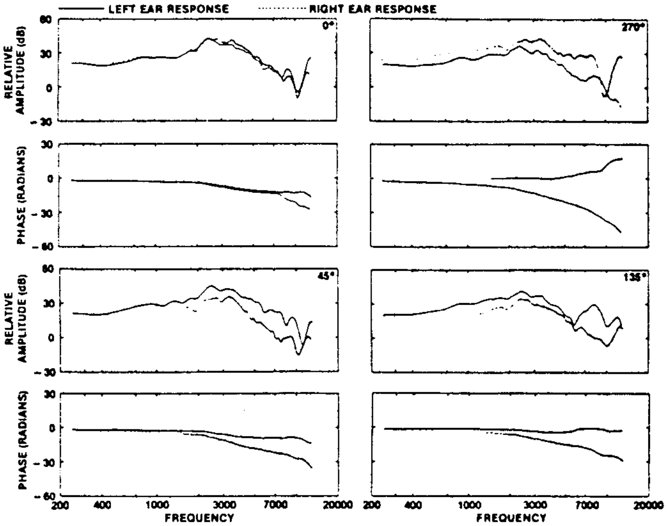


Figure 1. The eight panels plot HRTFs for a single listener for four source locations. (See text.)

A table of "location filters" is constructed from the pairs of HRTFs by first transforming them to the frequency domain, removing the spectral effects of the original loudspeakers and headphones using Fast Fourier Transform techniques, and then transforming back to the time domain. The table of corrected HRTFs acts as input to a realtime signal-processing device designed by Scott Foster and currently being prototyped at Ames. The device performs a mathematical convolution of an arbitrary signal, such as a voice, with the filter coefficients thus "placing" the signal within the perceptual 3-space of the user. The initial configuration allows up to three independent and simultaneous sources, requiring more than 100,000,000 multiply-accumulates per second. The resulting data stream is converted to an analog signal and presented over headphones.

Motion trajectories and static locations at greater resolution than the empirical measurements are simulated by interpolation. When integrated into VIEW, the operator's position can be monitored in realtime so that several simultaneous sources can be maintained in fixed locations or motion trajectories relative to the user. Such head-coupling should substantially enhance the simulation since previous work indicates that head movements may be important for localization (Thurlow & Runge, 1967). Informal tests at the University of Wisconsin and at Ames also suggest that the approach is feasible; simple linear interpolations between locations as far apart as 60 degrees azimuth are perceptually indistinguishable from stimuli synthesized from measured coefficients. As with any system required to compute data "on the fly", the term "real-time" is a relative one. The present system is designed to have a maximal latency or directional sampling interval of 30-50 msec. Recent work on the perception of auditory motion by Perrott (1982) and others indicates that this value should be acceptable since it is less than the minimum perceivable delay measured by the Minimum Audible Movement Angle for speeds of up to 360 degrees/sec.

PSYCHOPHYSICAL VALIDATION

Basic Synthesis Technique

The working assumption of the synthesis technique is that if, using headphones, we could produce ear canal waveforms identical to those produced by a free-field source, we would duplicate the free-field experience. The only conclusive test of this assumption must come from psychophysical studies in which free-field and synthesized, free-field listening are directly compared. A recent study at the University of Wisconsin has confirmed the perceptual adequacy of the basic approach for static sources. The stimuli were similar to those used to measure the HRTFs and subjects were blindfolded. Table 1 compares six subjects' absolute judgements of location under free-field and synthesized conditions. Note that overall goodness of fit between actual and estimated source coordinates are 0.89 or better for the synthesized stimuli and 0.91 or better for free-field sources. The correlation measures indicate that while source azimuth appears to be synthesized nearly perfectly, synthesis of source elevation is less than perfect, particularly for subjects that have difficulty judging elevation in the free-field. Similarly, the percentage of front-back reversals for synthesized stimuli is greater for all listeners.

Localization performance was also examined for regional differences, e.g., low, middle, and high elevations crossed with front, side, and back azimuths. For the sake of brevity, the data are only summarized here. As in previous free-field studies (e.g., Oldfield & Parker, 1984), localization is generally poorest at high elevations in the rear for both free-field and synthesized stimuli. Like the global data, the nine regions show a close correspondence between each subject's judgements in the free-field and synthesized conditions, although correlations are somewhat lower and front-back confusions approximately double for synthesized stimuli. Again, performance is generally better for azimuth than elevation, but with no obvious regional differences in the size and variability of errors for free-field and headphone listening. Thus, while individual differences do occur, the pattern of results across stimulus conditions is quite consistent for a given subject. That is, the frequently cited notion of "good" and "bad" localizers (Butler & Belendiuk, 1977) appears to be supported by both the global and regional analyses.

Global Measures of Localization Performance

Subject ID	Goodness of Fit	Azimuth Correlation	Elevation Correlation	% Front-Back Reversals
SDE	**.91** (.89)	**.993** (.992)	**.69** (.50)	**13.4** (20.6)
SDH	**.95** (.95)	**.991** (.983)	**.94** (.86)	**4.9** (10.3)
SDL	**.97** (.95)	**.995** (.991)	**.92** (.91)	**6.9** (13.6)
SDM	**.98** (.96)	**.995** (.996)	**.95** (.94)	**4.6** (9.4)
SDO	**.97** (.97)	**.996** (.996)	**.97** (.95)	**3.0** (14.7)
SDP	**.99** (.97)	**.999** (.997)	**.98** (.92)	**3.7** (5.6)

Table 1. Measures of localization performance for 6 subjects are compared for free-field (boldface type) and synthesized stimuli (in parentheses).

Acoustic Determinants of Performance

The presence of substantial individual differences in localization behavior suggests that there are acoustic features peculiar to each subject's HRTFs which influence performance. Thus, the use of "average" transforms, or even measurements derived from normative manikins such as the Kemar, may not be an optimum approach for simulating free-field sounds. For example, Figure 2 illustrates the between-subjects variability in the magnitude responses for a single source location. It can be seen that any straightforward averaging of these functions would tend to smooth the peaks and valleys, thus removing potentially significant features in the acoustic transforms.

Alternatively, it may be possible to enhance or avoid specific features of transforms which result in good or bad localization. The psychophysical data indicate that elevation is particularly difficult to judge, especially for subject SDE. A preliminary analysis of elevation coding suggests that there is an acoustic basis for this poor performance. Figure 3 plots "interaural elevation dependency" functions for four subjects' magnitude data. Interaural difference functions were first computed by dividing all leading ear HRTFs by their corresponding lagging ear HRTFs. These difference functions were then "normalized" to zero elevation by dividing all the difference functions at a given azimuth by the difference function at zero elevation and that azimuth. Finally, the normalized, elevation dependency functions were averaged over all azimuths. Thus, for each subject in Figure 3, the six functions show how interaural intensity changes for the different elevations with respect to zero elevation (the flat function). In spite of the large intersubject variabilty exemplified by Figure 2, the functions for the better localizers, SDL, SDH, and SDO, are quite similar to each other and show clear elevation dependencies. (The functions for the two subjects not shown were also similar.) SDE's, on the other hand, are different from the other subjects and show little elevation dependency. Thus, it appears that SDE's poor performance in judging elevation for both real and synthesized stimuli may be due to the lack of distinctive acoustic features in the HRTFs.

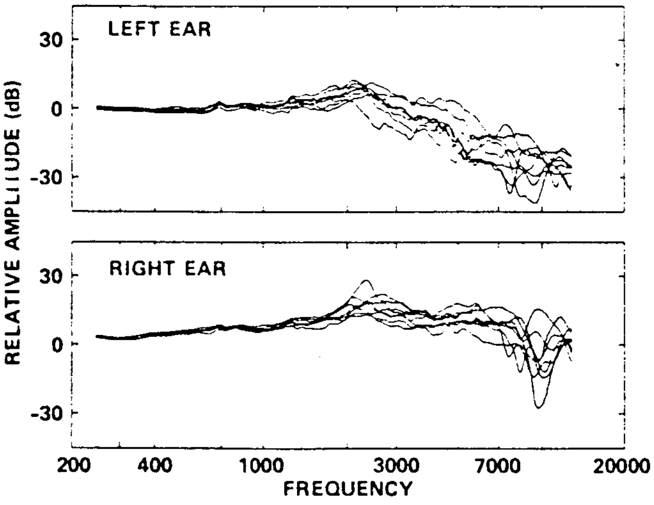


Figure 2. Magnitude responses for a single source position for 8 subjects. The left and right ears are plotted separately.

HUMAN FACTORS APPLICATIONS

There are many potential applications of a 3-D acoustic display. Such an interface will be valuable in any context where the user's awareness of his spatial surroundings is important. This information is particularly critical under high workload, especially when visual cues are degraded or absent and direct sensation of the auditory world may not be possible or desirable. Examples include tracking the locations of other objects and traffic during extra-vehicular activity in space or Nap-of-the-Earth flight in the helicopter, acoustic navigation displays for the blind, advanced teleconferencing environments, and sophisticated auditory image processing for the entertainment industry (also see Fisher, this volume).

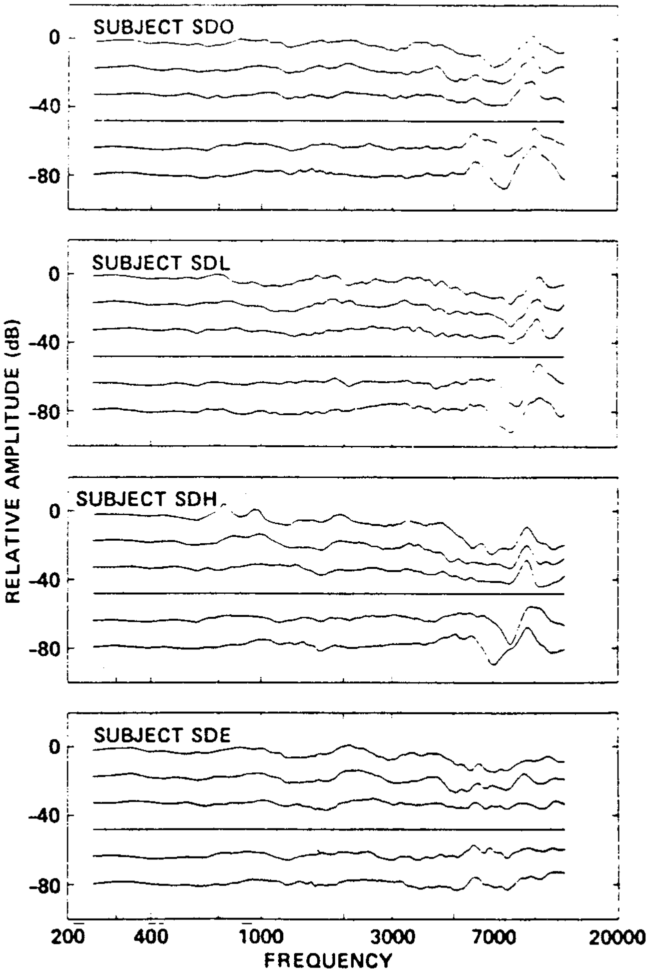


Figure 3. Interaural elevation dependency functions are plotted for 4 subjects. From top to bottom, the functions within a panel represent elevations of 54, 36, 18, 0 (the reference elevation), -18, and -36 degrees (See text.)

One application suggested for the space station is the use of a spatial auditory display during proximity operations for monitoring traffic in the vicinity of the station and controlling docking with the shuttle (Figure 4). Localized acoustic cues can provide direct information about spatial relationships in a situation with limited field-of-view and no natural acoustic input. The operator need not rely solely on his/her interpretation of a three-dimensional space which has been transformed into a two-dimensional visual display or camera viewpoint. Even quite simple auditory cues, such as a sound signalling direction, distance, and finally contact with the virtual shuttle could greatly aid in supervising the remote docking operation.

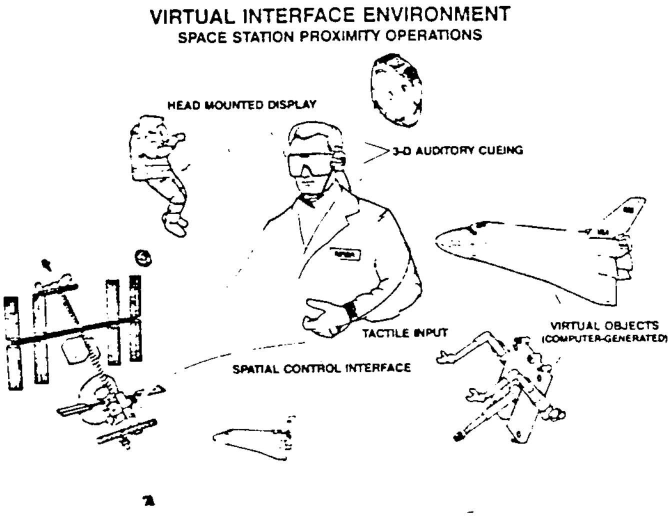

VIRTUAL INTERFACE ENVIRONMENT
SPACE STATION PROXIMITY OPERATIONS
HEAD MOUNTED DISPLAY
3-D AUDITORY CUEING
TACTILE INPUT
VIRTUAL OBJECTS (COMPUTER-GENERATED)
SPATIAL CONTROL INTERFACE

Figure 4. A possible application of 3-D auditory cues: Proximity Operations Display for the Space Station.

CONCLUSIONS

The psychophysical data indicate that the technique provides a successful first attempt at synthesizing localized sound, particularly for source azimuth. However, many factors not directly addressed here will need to be considered in producing a veridical percept. For example, relatively little is known about the nature of auditory motion perception (Perrott; Strybel, this volume), the critical cues for distance (Coleman, 1962) or the potential contribution of reverberation (Blauert, 1983) to an accurate perception of auditory space. Earlier work also suggests that localization may be influenced by the nature of accompanying visual cues (Gardner, 1968), the listener´s familiarity with the sounds to be localized (Coleman, 1962), and individual differences in pinna cues (Butler and Belendiuk, 1977). Indeed, the data reported here suggest that differences in the pinnae determine an individual´s localization performance, especially for elevation.

Finding solutions to problems like listener-dependence will have a critical impact on the general utility of a three-dimensional auditory display in any application. In the past it has not been possible to adequately test many aspects of these questions simply because it was technically too difficult to put the stimuli under direct experimental control. The realtime signal-processing device under development at Ames should prove a very useful tool for examining some of these issues as well as furnish the basic technology for sophisticated acoustic displays.

REFERENCES

Blauert, J. (1983) Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press: Cambridge, MA.

Brooks, F.P. (1988) Grasping reality through illusion -- Interactive graphics serving science. Proc. CHI´88, ACM Conf. Hum. Fac. Comp. Sys., Washington, D.C., 1-11.

Butler, R.A. & Belendiuk, K. (1977) Spectral cues utilized in the localization of sound in the median sagittal plane. J. Acoust. Soc. Am., 61, 1264-1269.

Coleman, P.D. (1962) Failure to localize the source distance of an unfamiliar sound. J. Acoust. Soc. Am., 34, 345-346.

Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.

Furness, T.A. (1986) The super cockpit and its human factors challenges. Proc. Hum. Fac. Soc., 1986 (1), 48-52.

Gardner, M.B. (1968) Proximity image effect in sound localization. J. Acoust. Soc. Am., 43, 163.

Gardner, M.B. & Gardner, R.S. (1973) Problem of localization in the median plane: Effect of pinnae cavity occlusion. J. Acoust. Soc. Am., 53, 400-408.

Oldfield, S.R. & Parker, S.P.A. (1984) Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent. Perc., 13, 601-617.

Perrott, D.R (1982) Studies in the perception of auditory motion. In Localization of Sound: Theory and Applications, R.W. Gatehouse (Ed.), Amphora Press: Groton, CN, 169-193.

Plenge, G. (1974) On the difference between localization and lateralization. J. Acoust. Soc. Am., 56, 944-951.

Lord Rayleigh [Strutt, J.W.] (1907) On our perception of sound direction. Phil. Mag., 13, 214-232.

Shaw, E.A.G. (1974) The external ear. In Handbook of Sensory Physiology, Vol. V/1, Auditory System, W.D. Keidel & W.D. Neff (Eds.), Springer-Verlag: New York.

Sutherland, I.E. (1968) Head-mounted three-dimensional display. Proc. Fall Joint Comp. Conf., 33, 757-764.

Thurlow, W.R. & Runge, P.S. (1967) Effect of induced head movements on localization of direction of sounds. J. Acoust. Soc. Am., 42, 480-488.

VIRTUAL INTERFACE ENVIRONMENT WORKSTATIONS

S.S. Fisher, E.M. Wenzel, C.Coler, M.W. McGreevy
Aerospace Human Factors Research Division
NASA Ames Research Center
Moffett Field, California 94035

ABSTRACT

A head-mounted, wide-angle, stereoscopic display system controlled by operator position, voice and gesture has been developed at NASA's Ames Research Center for use as a multipurpose interface environment. This Virtual Interface Environment Workstation (VIEW) system provides a multisensory, interactive display environment in which a user can virtually explore a 360-degree synthesized or remotely sensed environment and can viscerally interact with its components. Primary applications of the system are in telerobotics, management of large-scale integrated information systems, and human factors research. System configuration, research scenarios, and research directions are described.

RESEARCH OBJECTIVES

The primary objective of this research is to develop a multipurpose, multimodal operator interface to facilitate natural interaction with complex operational tasks and to augment operator situational awareness of large-scale autonomous and semi-autonomous integrated systems. The system should specifically provide a uniform interface for multiple task supervision, be reconfigurable for varying levels of operator skill, training and preference, and have an operator interface configuration that features human matched displays and controls for transparent, natural system interaction and reduced training requirements.

In the Aerospace Human Factors Research Division of NASA's Ames Research Center, an interactive Virtual Interface Environment Workstation (VIEW) has been developed to aid in design, simulation and evaluation of advanced data display and management concepts for operator interface design. The VIEW system provides a virtual auditory and stereoscopic image surround that is responsive to inputs from the operator's position, voice and gestures. As a low-cost, multipurpose simulation device, this variable interface configuration allows an operator to virtually explore a 360-degree synthesized or remotely sensed environment and viscerally interact with its components.

Application areas of the virtual interface environment research are focused in:

1. Development of workstations for complex operational tasks such as telerobotic and telepresence control of remotely operated robotic devices and vehicles that require a sufficient quantity and quality of sensory feedback to approximate actual presence at the task site.

2. Development of concepts and guidelines for `portable' multi-modal information management systems such as an EVA Spacesuit visor display, with subsequent development of workstations for supervision and management of large-scale integrated information systems in which data manipulation, storage and retrieval, and system monitoring tasks can be spatially organized.

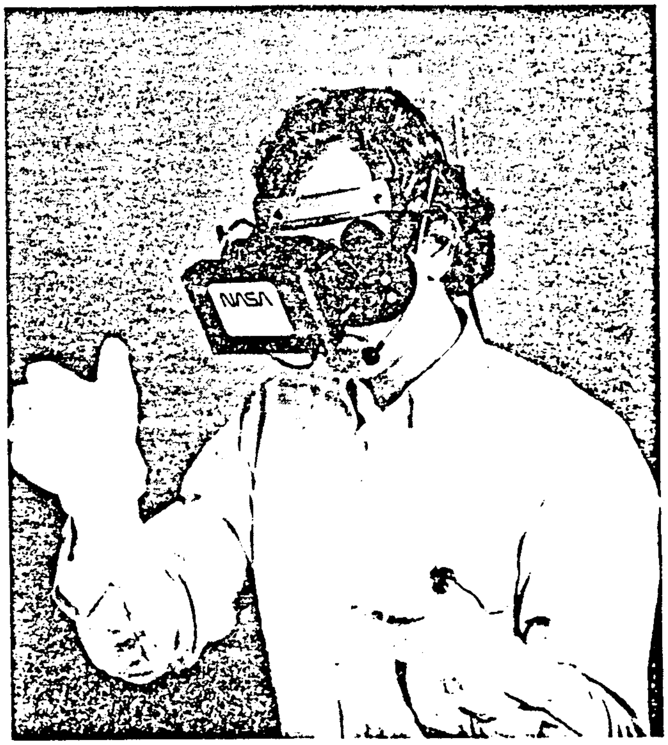


Figure 1. Head-coupled', stereoscopic display system with DataGlove tactile input technology for virtual object manipulation.

A more general research objective includes use of this display system to synthesize interactive test environments for aerospace human factors research in such areas as: spatial habitability research; rapid prototyping of display and workstation configurations; research on effective transfer of spatial information and visualization of multidimensional data; and spatial cognition research on multisensory integration.

SYSTEM CONFIGURATION

The current Virtual Interface Environment Workstation system consists of: a wide-angle stereoscopic display unit, glove-like devices for multiple degree-of-freedom tactile input, connected speech recognition technology, gesture tracking devices, 3D auditory display and speech-synthesis technology, and computer graphic and video image generation equipment. This hardware is integrated with a realtime Unix workstation that supports the computations required to drive an external high-performance realtime 3D graphics system, processes input from up to 12 realtime input peripherals (e.g., the trackers and gloves), and provides other realtime task processing. A collection of software called the `simulation framework´ has also been developed that consists of a well-documented library of functions to provide access to all of the system peripherals and software services, and of a collection of source files and simulation software that demonstrates the use of the major hardware and software components that make up the VIEW system in order to facilitate system reconfiguration for changing research requirements.

VIEW Visual Display Technology

The VIEW Display subsystem is an integrated technology package comprised of image display elements, optics, and electronics which together provide a wide angle, stereoscopic image environment that closely matches human binocular visual capabilities. When combined with high-resolution, magnetic 6 degree-of-freedom head and limb position tracking technology, the displayed imagery appears to completely surround the user in 3-space and provides interactive viewing and manipulation capabilities. Specific areas of display research in progress include: development of orthostereoscopic and remote presence display systems for improved operator performance in telerobotics and telepresence control; display of multi-dimensional, graphic representations of knowledge-based information systems to facilitate supervision and management of large-scale integrated systems; and 3D simulation of veridical experimental environments for human factors research.

Head-mounted displays. The head-coupled display unit uses two passive, twisted nematic, monochromatic, liquid crystal display screens presented to each eye of the user through wide-angle optics. Resolution of each display is 640x220 pixel elements with 320 distinguishable vertical lines and approximately 16 levels of grayscale. The LCD elements incorporate a diamond-shaped pixel geometry that tends to reduce the apparent fixed pattern spatial noise by effectively doubling the horizontal and vertical fixed pattern spatial frequencies. Currently, the overall size of each display is 3.2" diagonally in a 4:3 aspect ratio. The transmissive LCD screens are backlighted with miniature flourescent light sources that result in a contrast ratio of approximately 7:1. This LCD technology is capable of address times of 63.5 microseconds per line with typical pixel response times of approximately 32 milliseconds. Imagery displayed on the screens is a NTSC standard video signal generated by computer, remote video cameras, or a combination of these input media with other video sources such as optical video disk. Horizontal position of the images on the display screens can be electronically controlled by sliders on the display casing to adjust for image/optics alignment and for variable convergence requirements. The display unit is connected to support electronics that provide power and conditioned video signals through a 20 foot cable. In addition, these electronics provide stereo alignment patterns, user selection of biocular or binocular imagery output to the viewer, and can simultaneously output multiplexed stereo video signals for videotape storage and replay.

Optics. The display screens are viewed directly through a pair of wide angle magnifying lenses that are mounted approximately 5 mm from the screens with adjustments for variable accommodation requirements such as relative tilt and distance of LCD substrate from optics. These optics provide a 120 degree horizontal and vertical field of view for each eye and up to a 90 degree binocular field overlap. Total instantaneous visual field of view is approximately 120 degrees. The 2.75" diameter of each optical element requires a minimum 4" diagonal display size to completely fill the available field of view. With 640x220 pixel resolution, each pixel in the diagonal array (i.e. each distinguishable vertical line) subtends 0.38 degrees (22.5 minutes) of horizontal visual field. The wide angle optics create a pincushion distortion in the displayed imagery that requires a barrel distortion compensation in the image generation or image capture technology for correct scene representation.

Mounting configurations. The display and optics unit is positioned directly in front of the viewer's eyes and coupled to head motion by means of a lightweight headgear configuration. Adjustments are provided for varying eye-relief distance and size of the supporting head band. Eye-cups around the optic elements are included to reduce interference of ambient light. The entire unit is counterweighted to transfer the weight of the display and optics unit to a point directly over the operator's spine. Also included on the headgear unit are: a support for attaching the head position tracking sensor, a microphone for input to the connected speech recognition subsystem, and earphones for auditory display feedback to the operator. See Figure 1.

Desk-mounted viewer. The wideangle viewing optics package is also implemented in a movable arm-mounted, workstation configuration for evaluation in design and engineering applications such as the display of three dimensional graphic databases used in Computational Fluid Dynamics Research. This viewer incorporates CRT display technology with a display resolution of 400 x 400 pixels. The position and orientation of the display unit and mounting arm is transduced and provided to the host computer to allow the operator to change viewpoints within the database in real time.

Image generation and capture. The display subsystem is used to view Virtual Environments that are synthesized with 3D computer-generated imagery, or that are remotely sensed by user-controlled, stereoscopic video camera configurations. The computer image system enables high performance, realtime 3D graphics presentation at resolutions of 640x480 and 1,000 x 1,000 pixels. This imagery is generated at rates up to 30 frames per second as required to update image viewpoints in coordination with head and limb motion. Dual independent, synchronized display channels are implemented to present disparate imagery to each eye of the viewer for true stereoscopic depth cues. For realtime video input of remote environments, two miniature CCD video cameras are used to provide stereoscopic imagery. Development and evaluation of several head-coupled, remote camera platform and gimbal prototypes is in progress to determine optimal hardware and control configurations for remotely controlled, free-flying or telerobot mounted camera systems. Research efforts also include the development of realtime signal processing technology to combine multiple video sources with computer generated imagery.

VIEW Auditory Display Technology

Binaural auditory display. The initial auditory subsystem for VIEW is capable of presenting a wide variety of binaural sounds to the operator via headphones using sound-synthesis technology developed for music synthesizers. This system has been developed so that a basic sound library may be designed and stored off-line for later realtime manipulation using an interactive editing interface and/or `front-panel' capabilities of the synthesizer. Thus, different sound signatures (e.g., different timbres or temporal patterns) can be associated with different objects or types of information in the visual displays to achieve a quasi-semantic or symbolic capability. Then, for each of the independent and simultaneous sound signatures or `objects', a set of global sound parameters can be dynamically coordinated with the other display parameters via the interface between the host computer and the synthesizer. For example, stereo panning, pitch, and intensity can be associated with position, size, or distance of objects in the visual display. Also, commercially-available speech-synthesis technology with unlimited vocabulary is used to provide an additional display capability for voice report requirements and verbal acknowledgement of system input. This synthesizer is a text to speech converter that generates a human-sounding voice with capability to modify pronunciation of output words, to change speaking rate, and to select from six different voice qualities.

The primary function of this initial auditory display is to provide both discrete and dynamic auditory cues which can augment or supply information missing from the visual or gestural displays. For example, discrete cues might signal contact between telerobot end-effectors and target objects or alert the operator to attend to information in a data window which is currently out of the field-of-view. Similarly, auditory parameters can be dynamically modulated and coordinated with the other subsystem displays for such tasks as monitoring the relative positions of dual robot arms in order to keep them witnin a proscribed motion envelope.

3D Auditory display. Additional research is underway at Ames to develop an auditory display prototype that is capable of synthetically generating three-dimensional sound cues in realtime. These cues are presented via headphones and are perceived outside of the head at a discrete distance and direction in the 3-space surrounding the user. When integrated into the VIEW system, position of the operator can be monitored in realtime and the information used to maintain several localized sound cues in fixed positions or in motion trajectories relative to the user. This capability will further aid the operator's situational awareness by augmenting spatial information from the visual display and providing outside the field-of-view navigational and cueing aids (see Wenzel,et.al.,1988, this volume).

VIEW Interaction

Visceral interaction with the Virtual Environment surrounding the user is enabled by speech and gesture input technologies. Current research includes configuration and evaluation of these input modalities, both separately and in combination, to facilitate effective, natural user interaction in various simulated task environments.

Speech recognition. The VIEW system includes commercially-available, speaker-dependent, connected speech-recognition technology that allows the user to give system commands in a natural, conversational format that can not be achieved with highly constrained discrete word recognition systems or through keyboard input. Typical speech mediated interactions are requests for display/report of system status, instructions for supervisory control tasks, and verbal commands to change interface mode or configuration.

DataGloves. For tactile interaction with the displayed three dimensional virtual environment, the user wears lightweight glove-like devices that transmit data-records of arm, hand and finger shape and position to a host computer. The gloves are instrumented with fiber optic flex-sensing devices at each finger joint and between fingers. Electromagnetic tracking sensors like those described for sensing the operator's head coordinates are also mounted on each glove to transmit relative position and orientation of the hands and arms to the host system. The system also includes preprocessor resident subroutine and editing capability to recognize and transmit specific hand/finger gestures for use in supervisory control modes, and software capability to quickly calibrate system measurements for use with different operator hand sizes. Primary uses of this technology are to provide a three-dimensional cursor in the displayed environment, and to enable interaction with virtual or remote objects. Current implementations of this research at Ames include a three-dimensional graphic database of an articulated hand that, in the virtual display environment, appears spatially correspondent with the viewer's real hand and is directly controlled by the instrumented glove device. With this capability, the operator can pick-up and manipulate virtual objects that appear in the surrounding virtual environment. Similarly, in VIEW system developed data-management environments, multiple windows of information and simulated control panels are positioned, sized and activated by simply manipulating the virtual objects in 3-space. In coordination with the connected speech recognition technology, the hand and arm gesture information can also be used to effect and/or direct speech-indicated actions in the synthesized or remote environment (e.g., control of robotic arms and end-effectors, and associated control of remote camera viewpoints) (Fisher, 1986; Foley, 1987).

RESEARCH SCENARIOS

Telerobotics

Control of autonomous and semi-autonomous telerobotic devices and vehicles requires an interface configuration that allows variable modes of operator interaction ranging from high-level, supervisory control of multiple independent systems to highly interactive, kinaesthetic coupling between operator and remote system. An appropriate interface for supervisory control modes will provide the operator with multiple viewpoints of the remote task environment in a multi-modal display format that can be easily distributed and reconfigured according to changing task priorities. For remote operations that cannot be performed autonomously, the interface will need capability to quickly switch to interactive control. In this telepresence mode, the operator will require a sufficient quantity and quality of sensory feedback to approximate actual presence at the remote task site.

The virtual environment display system is currently used to interact with a simulated telerobotic task environment. The system operator can call up multiple images of the remote task environment that represent viewpoints from free-flying or telerobot-mounted camera platforms. Three-dimensional sound cues give distance and direction information for proximate objects and events. Switching to telepresence control mode, the operator's wide-angle, stereoscopic display is directly linked to the telerobot 3D camera system for precise viewpoint control. Using the tactile input glove technology and speech commands, the operator directly controls the robot arm and dexterous end effector which appear to be spatially correspondent with his own arm. Current experimental research includes evaluation of operator performance in teleoperation placement tasks and of supervisory control interface configurations.

Dataspace

Similar to interface requirements for supervisory control of telerobotic systems, the operator interface for large-scale integrated information systems such as Space Station will also require a range of interaction modalities for operator intervention in the case of system degradation or conflicts in resource allocation. Efficient supervision of these automated systems will depend on highly graphic, multi-dimensional status representations of the numerous sub-systems in a format that can be easily monitored in parallel with other mission related tasks such as planning and scheduling, telerobot supervision, and communication activities. In the event of system conflict or malfunction, the interface environment should enable natural interaction with multi-modal, knowledge-based warning and advisory systems.

Advanced data display and management concepts for this task environment are being developed with the virtual environment display technology. Current investigations include use of the system to create a display environment in which data manipulation and system monitoring tasks are organized in virtual display space around the operator. Through speech and gesture interaction with the virtual display, the operator can rapidly call up or delete information windows and reposition them in 3-space. Three-dimensional sound cues and speech-synthesis technologies are used to enhance the operators overall situational awareness of the virtual data environment. The system also has the capability to display reconfigurable, virtual control panels that respond to glove-like tactile input devices worn by the operator. Major research issues include development of multi-modal data access and manipulation guidelines; concepts for multi-dimensional, graphic representation of knowledge-based information systems; and definition of interface configurations for shared workspace environments in collaborative systems management.

DISCUSSION

Research Directions

Projected near term technology developments for the VIEW system include improved and distortion-corrected computer generated imagery processed at significantly higher frame rates; High resolution, color display elements; Integration of 3D auditory display technology for spatially correspondent, synthesized sound cueing; Tactile feedback capability; Multiple viewpoint display integration; And multiple VIEW display stations for shared workspace configurations in which each user is graphically present and interactive in the Virtual Environment.

Conclusions

Unlike most technology for 360-degree visual simulation environments, the virtual environment display system does not make use of large, expensive, special purpose projection configurations. The described system is portable and low-cost without large space and equipment requirements. In comparison to other research efforts in head-mounted displays, this system is unique in presenting a stereoscopic image that closely matches human binocular vision capabilities and in its configuration with state-of-the-art auditory, speech and tactile input technology.

The VIEW systems' capabilities for providing a highly graphical, uniform interface for varying task environments and level of interaction may reduce operator workload and training requirements and increase productivity, and it's use of muti-modal input channels and augmentation of visual displays to provide unobtrusive redundancy may increase accuracy and efficiency in human-machine interaction. Also, multiple viewpoint presentation of task related information in a 360-degree, dynamic stereoscopic display appears to increase situational awareness and effectiveness in monitoring cognitively demanding spatial tasks such as construction and proximity operations or spatial representations of knowledge-based systems and activities.

REFERENCES

Fisher, S.S. (1986) Telepresence Master Glove Controller for Dexterous Robotic End-Effectors, Advances in Intelligent Robotics Systems, D.P.Casasent, Editor, Proc. SPIE 726, 1986.

Foley, J.D. (1987), Interfaces for Advanced Computing. Scientific American, 257(4), 126-135.

Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988) A Virtual Acoustic Display for Conveying Three-Dimensional Information. Proc. Hum. Fac. Soc., 1988.

- **Hide:** Stop a primitive from being available to the user.

- **Set Contents:** Either cause a parent-child relationship to exist if this is a context primitive, or set the contents of the Text/Marker/Symbol primitive.

- **Set Attributes:** Primitives may have some presentation specific details which do not really contribute to the information, but may be aesthetically important.

Example

One common situation which could be used as an example to help explain this strategy is an application program that needs to present a pop-up menu to the user.

To accomplish this, the application would produce output primitive operators to guide appropriate modifications to the primitive feedback tree. A result of this would be the addition of a new subtree as shown in Figure 1.

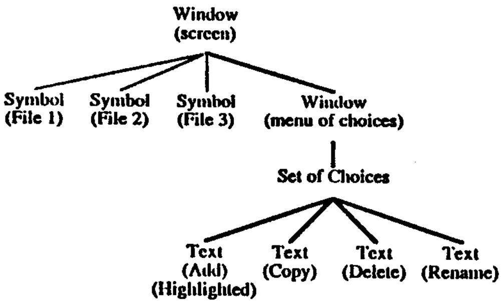


Figure 1: A new subtree is added

This primitive feedback tree would then be interpreted visually as in Figure 2, or auditorially as in Figure 3, depending on the interface manager chosen for use by the end user.

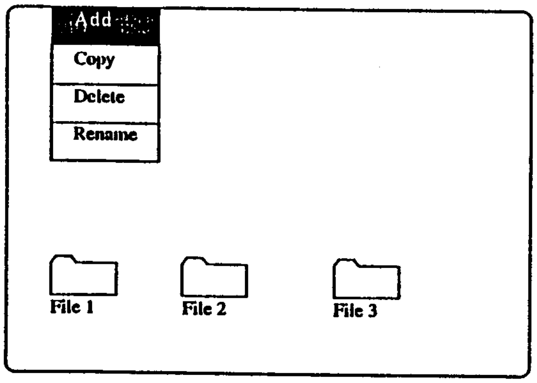
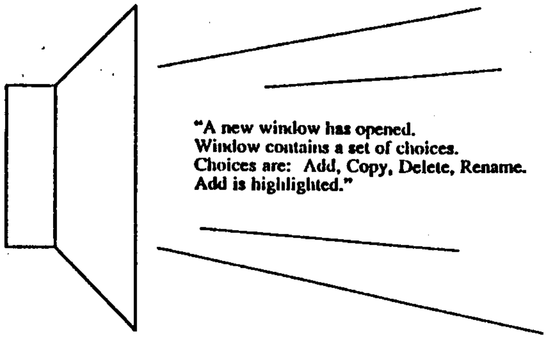


Figure 2: Visual presentation of the feedback tree



"A new window has opened.
Window contains a set of choices.
Choices are: Add, Copy, Delete, Rename.
Add is highlighted."

Figure 3: Auditory presentation of the feedback tree

Summary

A strategy has been proposed to allow people with disabilities the choice of feedback presentation modality. To accomplish this a set of output primitives has been developed which presents the content, context and relationships between information in an unambiguous tree structure. This structure is then interpreted by an interface manager to determine the final form of the presentation.

Acknowledgements

This project was funded by IBM Corporation as a Shared University Research Project with the University of Toronto and the Hugh MacMillan Medical Centre, Toronto, Canada.

DEVELOPMENT OF A THREE-DIMENSIONAL AUDITORY DISPLAY SYSTEM

ELIZABETH M. WENZEL, FREDERIC L. WIGHTMAN, SCOTT H. FOSTER

Abstract

We propose that the most powerful method of auditory cueing takes direct advantage of human perceptual capabilities, providing a dynamic, multidimensional pattern of events which conveys meaning about objects in the spatial world. Applications of such a three-dimensional auditory display involve any context in which the user's situational awareness is critical, particularly when visual cues are limited or absent. Examples include air traffic control displays, advanced teleconferencing environments, and monitoring telerobotic activities in hazardous situations.

This type of display system requires the ability to generate localized sound cues in a flexible and dynamic manner. Whereas this can be achieved with an array of real sound sources or loudspeakers, the prototype device being developed at NASA-Ames maximizes flexibility and portability by synthetically generating three-dimensional sound cues in realtime for delivery through headphones. Unlike conventional stereo, sources can be perceived outside the head at discrete distances and directions from the listener. When completed, the device will be integrated with the Virtual Interactive Environment Workstation (VIEW), a head-mounted, wide-angle, stereoscopic display system controlled by operator position, voice, and gesture (Fisher, et. al., 1986).

Previous research in psychoacoustics suggests that perceptually-veridical localization over headphones is possible if both the direction-dependent pinna cues and the more well understood cues of interaural time and intensity are adequately synthesized. Although the realtime device is not yet finished, recent studies at the University of Wisconsin have confirmed the perceptual adequacy of the basic approach to synthesis.

Synthesis: Theory & Technique

The 'Duplex Theory' of human sound localization, first described by Lord Rayleigh (1907), states that the two cues primarily responsible for localization are interaural differences in time of arrival for low-frequency waveforms and interaural differences in intensity at high frequencies. Although the theory has long dominated thinking about binaural hearing, recent research points to serious limitations with this formulation. For example, it cannot account for the ability of subjects to precisely locate sounds on the median plane where interaural cues are minimal or absent (see Blauert, 1983). Similarly, when subjects listen to stimuli over headphones, they are perceived as inside the head even though interaural temporal and intensity differences appropriate to an external source location are introduced. A variety of research indicates that deficiencies of the Duplex Theory stem from an inadequate consideration of filtering by the outer ears or pinnae; Shaw (1974) demonstrated that spectral shaping by the pinnae is indeed direction-dependent, while other experiments showed that the absence of pinna cues degrades localization accuracy (Gardner & Gardner, 1973; Oldfield & Parker, 1984), and Plenge (1974) concluded that pinna cues are primarily responsible for externalization or the sensation of 'out-there-ness'. Such studies suggest that perceptually-veridical localization over headphones is possible if both the direction-dependent pinna cues and the more well understood cues of interaural time and intensity can be adequately synthesized.

A number of experimenters have been interested in simulating various features of auditory spatial perception. For example, one approach following from the work of Batteau (1968) simulates elevation with moderate success using a time-domain or comb-filtering technique (Watkins, 1978). Spatial processing has also been of interest in computer music where there is an emphasis on simulating auditory 'spaciousness', rather than veridical localization *per se,*,with such cues as reverberation, distance, and Doppler shifts (e.g., Moore, 1983). Other attempts have centered on synthesizing essentially all available cues by perfecting methods of binaural recording using both stationary (see Blauert, 1983, p. 358) and head-tracked stereophony (Doll, et. al., 1986). Similarly, much of the recent work in this area has been on developing accurate techniques for modelling the pinnae (Shaw, 1974) or measuring pinna transform functions (Mehrgardt & Mellert, 1977; Wightman & Kistler, 1980; Foster, 1986).

Our technique derives from the latter approach, preserving the frequency-dependent interaural time and intensity information via precise measurement of the outer ear filtering characteristics. Figure 1 illustrates the method.

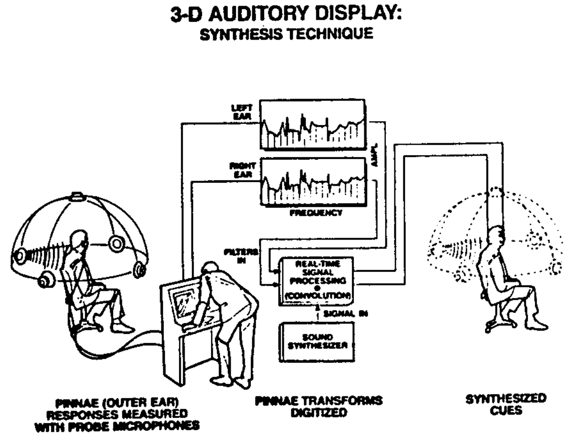

3-D AUDITORY DISPLAY:
SYNTHESIS TECHNIQUE
LEFT EAR
RIGHT EAR
AMPL.
FREQUENCY
FILTERS IN
REAL-TIME SIGNAL PROCESSING (CONVOLUTION)
SIGNAL IN
SOUND SYNTHESIZER
PINNAE (OUTER EAR) RESPONSES MEASURED WITH PROBE MICROPHONES
PINNAE TRANSFORMS DIGITIZED
SYNTHESIZED CUES

Figure 1. A conceptual illustration of the synthesis technique.

Probe microphones are placed near the two eardrums of a listener seated in the anechoic chamber of the Psychoacoustics Lab at Wisconsin. The subject is at the center of a spherical array of loudspeakers with a radius of 1.4 meters and a density of 15 degrees azimuth (range: 0 to 360 degrees) and 18 degrees elevation (range: -36 to 90 degrees where 0 is at ear level). The stimulus, consisting of a 4 Hz train of 100 acoustic clicks at a level of approximately 80 dB SPL, is then presented to the subject over one of the speakers. The responses of the probe microphones to the 100 clicks are averaged, digitized at a sampling rate of 50 kHz, and stored as 16-bit numbers. A new pair of filters or pinna transforms is then measured for each location.

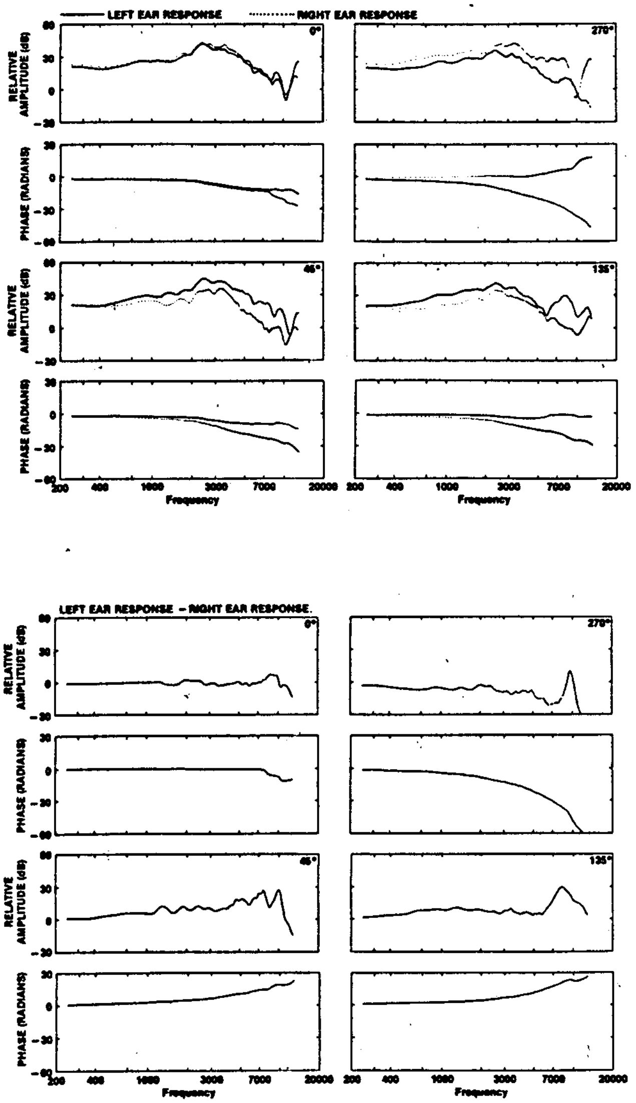


Figure 2. The panels on the top show pinna transform pairs for a single listener for four source locations. The bottom panels plot the interaural difference cues for the same locations. (See text.)

Figure 2 gives examples of empirical measurements of pairs of pinna transforms for a single listener for source locations at 0, 45, 135, and 270 degrees on the azimuthal plane (0 degrees is defined as directly in front, degrees increase counter-clockwise). The eight panels on the top plot the amplitude (relative decibels) and phase (radians) responses of the Fourier transforms for both the left (solid lines) and right ears (dotted lines) at the four locations. The eight bottom panels illustrate how the interaural difference cues change with location by plotting the difference between the left and right amplitude and phase responses as a function of frequency. Positive values indicate that the more intense and leading signal.

The table of 'location filters' derived from a complete set of measurements acts as input to a realtime signal-processing device designed by Scott Foster and currently being prototyped at Ames. The device convolves an arbitrary signal, such as a voice, with the measured filtered pairs, thus 'placing' it within the perceptual 3-space of the user. The initial configuration allows up to three independent and simultaneous sources, requiring approximately 100,000,000 multiply-accumulates per second. The resulting data stream is converted to an analog signal and presented via headphones. Prior to the actual synthesis, however, the spectral effects of the original loudspeakers and the headphones will have been removed digitally.

Motion trajectories and static locations at greater resolutions than the empirical measurements are simulated by interpolation. When integrated into VIEW, the operator's position can be monitored in realtime so that several simultaneous sources can be maintained in fixed positions or motion trajectories relative to the user. Such head-coupling should substantially enhance the simulation since previous work indicates that head movements may be important for localization (Thurlow & Runge, 1967). Informal tests at the University of Wisconsin and at Ames also suggest that this approach is feasible; simple linear interpolations between locations as far apart as 60 degrees azimuth are perceptually indistinguishable from stimuli synthesized from the measured coefficients. As with any system required to compute data 'on the fly', the term realtime is a relative one. The present system is designed to have a maximal latency or directional sampling interval of 50 msec. Recent work on the perception of auditory motion by Perrott (1982) and others indicates that this value should be acceptable since it is less than the minimum perceivable delay measured by the Minimum Audible Movement Angle for speeds of up to 360 degrees/sec.

Perceptual Verification of the Synthesis

A recent study at the University of Wisconsin has confirmed the perceptual adequacy of the basic synthesis technique. Table 1 compares the outcome of several global statistics for six subjects' absolute judgements of location under free-field (boldface type) and synthesized (in parentheses) conditions. For example, overall goodness of fit between actual and estimated source coordinates are 0.89 or better for the synthesized stimuli and 0.91 or better for real sources. The correlation measures, however, indicate that while source azimuth appears to be synthesized nearly perfectly for all listeners, synthesis of source elevation is less than perfect for some subjects. Similarly, the percentage of front-back reversals for synthesized stimuli is consistently greater for the same listeners. Thus, while individual differences do occur, the pattern of results across stimulus conditions is consistent within subjects. That is, the frequently cited notion of 'good' and 'bad' localizers appears to be supported by these data.

Global Measures of Localization Performance
Comparison of Free-Field vs. Synthesized Stimuli

Subject ID	Goodness of Fit	Azimuth Correlation	Elevation Correlation	% Front-Back Reversals
SDE	**.91** (.89)	**.993** (.992)	**.69** (.50)	**13.4** (20.6)
SDH	**.95** (.95)	**.991** (.983)	**.94** (.86)	**4.9** (10.3)
SDL	**.97** (.95)	**.995** (.991)	**.92** (.91)	**6.9** (13.6)
SDM	**.98** (.96)	**.995** (.996)	**.95** (.94)	**4.6** (9.4)
SDO	**.97** (.97)	**.996** (.996)	**.97** (.95)	**3.0** (14.7)
SDP	**.99** (.97)	**.999** (.997)	**.98** (.92)	**3.7** (5.6)

Table 1. Measures of free-field performance are in boldface type and measures of performance during synthesized conditions are in parentheses.

Human Factors Applications

The prototype auditory display being developed at the Aerospace Human Factors Research Division of NASA-Ames Research Center will form part of the Virtual Interactive Environment Workstation (VIEW), a head-mounted, wide-angle, stereoscopic display system controlled by operator position, voice, and gesture.

The potential applications of a 3-D acoustic display are many. Such an interface will be valuable in any situation where the user's awareness of his spatial surroundings is important. This information is particularly critical under high workload, especially when visual cues are degraded or absent and direct sensation of the auditory world may not be possible or desirable.

One application suggested for the Space Station is the use of a spatial auditory display during Proximity Operations for monitoring traffic in the vicinity of the station and controlling docking with the shuttle (Figure 3). Localized acoustic cues can provide direct information about spatial positions and relationships in a situation with limited field-of-view. The operator need not rely solely on his/her interpretation of a three-dimensional space which has been transformed into a two-dimensional visual display.

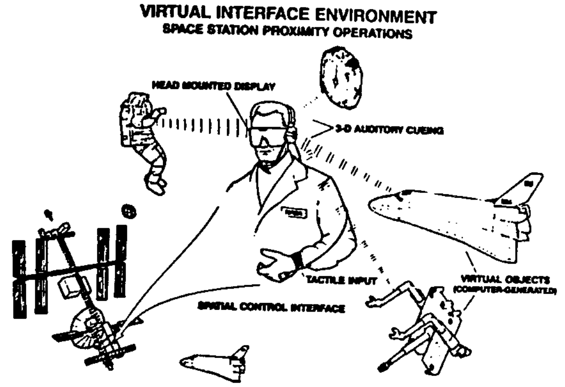

VIRTUAL INTERFACE ENVIRONMENT
SPACE STATION PROXIMITY OPERATIONS
HEAD MOUNTED DISPLAY
3-D AUDITORY CUEING
TACTILE INPUT
VIRTUAL OBJECTS (COMPUTER-GENERATED)
SPATIAL CONTROL INTERFACE

Figure 3. A possible application of 3-D auditory cues: Proximity Operations Display for the Space Station.

When integrated with VIEW, the 3-D auditory display will be coupled to the visual and gestural subsystems via realtime position-tracking so that simultaneous sound sources can be maintained in fixed positions or motion trajectories relative to the user. Thus, the display can provide an additional information channel while substantially augmenting the operator's situational awareness during activities such as supervision and control of telerobotic operations in hazardous environments (Figure 4). Even quite simple auditory cues, such as a sound signalling direction, distance, and finally contact with a virtual object could greatly aid the user in manipulating the remote world.

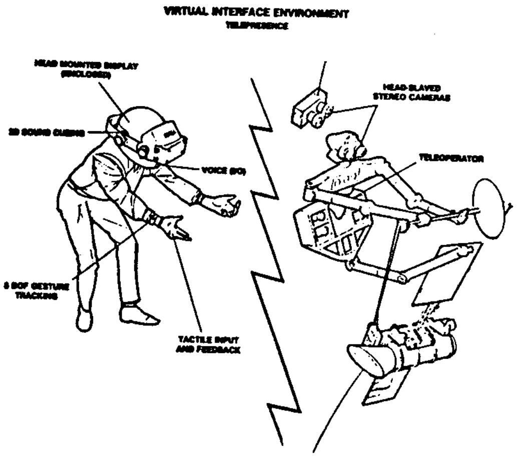

VIRTUAL INTERFACE ENVIRONMENT
TELEPRESENCE
HEAD MOUNTED DISPLAY (ENCLOSED)
3D SOUND CUEING
VOICE (I/O)
6 DOF GESTURE TRACKING
TACTILE INPUT AND FEEDBACK
HEAD-SLAVED STEREO CAMERAS
TELEOPERATOR

Figure 4. 3-D auditory displays can provide an additional information channel and augment situational awareness during telerobotic operations.

Future Research Directions

The initial psychophysical data indicate that the technique provides a successful first attempt at synthesizing localized sound, particularly for source azimuth. However, many factors not directly addressed here will need to be considered in producing a veridical percept. Earlier work, for example, suggests that localization may be influenced by the nature of accompanying visual cues (Gardner, 1968), the listener's familiarity with the sounds to be localized (Coleman, 1962), and individual differences in pinna transforms (Butler and Belendiuk, 1977). Also, relatively little is known about the nature of the critical cues for distance (Coleman, 1962) or the potential contribution of reverberation (Blauert, 1983) to an accurate perception of auditory space. In the past it has not been possible to adequately test many aspects of these problems simply because it was technically too difficult to put the stimuli under direct experimental control. The realtime signal-processing device under development at Ames should prove a very useful tool for examining some of these issues as well as furnish the basic technology for sophisticated auditory displays.

References

Batteau, D.W. (1968) Listening with the naked ear. In *The Neuropsychology of Spatially Oriented Behavior*, S.J. Freedman (Ed.), Dorsey Press, Homewood, IL.

Blauert, J. (1983) *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press: Cambridge, MA.

Butler, R.A. & Belendiuk, K. (1977) Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Am.*, 61, 1264-1269.

Coleman, P.D. (1962) Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. Am.*, 34, 345-346.

Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.

Fisher, S.S., McGreevy, M., Humphries, J., & Robinett, W. (1986) Virtual Environment Display System. *ACM Workshop Interact. 3D Graph.*, Chapel Hill, N.C.

Foster, S.H. (1986) Impulse response measurement using Golay codes. *Proc. Int. Conf. Acoust. Speech. Sig. Proc.*

Gardner, M.B. (1968) Proximity image effect in sound localization. *J. Acoust. Soc. Am.*, 43, 163.

Gardner, M.B. & Gardner, R.S. (1973) Problem of localization in the median plane: Effect of pinnae cavity occlusion. *J. Acoust. Soc. Am.*, 53, 400-408.

Mehrgardt, S. & Mellert, V. (1977) Transformation characteristics of the external human ear. *J. Acoust. Soc. Am.*, 61, 1567-1576.

Moore, F.R. (1983) A general model for spatial processing of sounds. *Comp. Mus. J.*, 7, 6-15.

Oldfield, S.R. & Parker, S.P.A. (1984) Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent. *Perc.,* 13, 601-617.

Perrott, D.R (1982) Studies in the perception of auditory motion. In *Localization of Sound: Theory and Applications,* R.W. Gatehouse (Ed.), Amphora Press: Groton, CN, 169-193.

Plenge, G. (1974) On the difference between localization and lateralization. *J. Acoust. Soc. Am.,* 56, 944-951.

Lord Rayleigh [Strutt, J.W.] (1907) On our perception of sound direction. *Phil. Mag.,* 13, 214-232.

Shaw, E.A.G. (1974) The external ear. In *Handbook of Sensory Physiology, Vol. V/1, Auditory System,* W.D. Keidel & W.D. Neff (Eds.), Springer-Verlag: New York.

Thurlow, W.R. & Runge, P.S. (1967) Effect of induced head movements on localization of direction of sounds. *J. Acoust. Soc. Am.,* 42, 480-488.

Watkins, A.J. (1978) Psychoacoustical aspects of synthesized vertical locale cues. *J. Acoust. Soc. Am.,* 63, 1152-1165.

Wightman, F.L., & Kistler, D.L. (1980) A New "Look" at Auditory Space Perception. In *Psychophysical, Physiological, and Behavioural Studies in Hearing,* G. van den Brink & F.A. Bilsen (Eds.), Delft University Press, The Netherlands.

About the Authors

Elizabeth M. Wenzel is a Research Psychologist in the Aerospace Human Factors Research Division at NASA-Ames Research Center. She received her Ph.D. in Cognitive Psychology, with an emphasis in auditory perception, from U.C., Berkeley in 1984.

Frederic L. Wightman is a professor in the Department of Psychology at the University of Wisconsin, Madison. He also holds a joint research appointment in psychoacoustics at the Waisman Center.

Scott H. Foster is an electrical engineer with over 10 years of experience in the design of audio signal processing hardware and software. He received his B.S. in Mathematics from M.I.T. in 1976.